B. A. Vorontsov-Vel'yaminov and others

PHYSICS OF STARS AND STELLAR SYSTEMS

Volume 2 of
A Course in Astrophysics and Stellar Astronomy

Edited by A. A. Mikhailov

Translated from Russian

Published for the National Aeronautics and Space Administration
and the National Science Foundation, Washington, D.C.
by the Israel Program for Scientific Translations

B. A. VORONTSOV-VEL'YAMINOV, T. A. AG.
A. N. DEICH, B. G. GORBATSKII, V. A. KRAT,
O. A. MEL'NIKOV, and V. V. SOBOLEV

PHYSICS OF STARS AND STELLAR SYSTEMS

Volume 2 of

A COURSE IN ASTROPHYSICS AND STELLAR ASTRONOMY

(Kurs astrofiziki i zvezdnoi astronomii)

Edited by A. A. MIKHAILOV

Editorial Board: A. N. DEICH, V. A. KRAT,
O. A. MEL'NIKOV, and V. V. SOBOLEV

Izdatel'stvo "Nauka"
Moskva 1962

Translated from Russian

Israel Program for Scientific Translations
Jerusalem 1969

NASA TT F-506
TT 68-50307

Published Pursuant to an Agreement with
THE NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, U.S.A.
and
THE NATIONAL SCIENCE FOUNDATION, WASHINGTON, D.C.

Copyright © 1969
Israel Program for Scientific Translations Ltd.
IPST Cat. No. 5139

Translated by Z. Lerman

Printed in Jerusalem by IPST Press
Binding: Wiener Bindery Ltd., Jerusalem

Available from the
U. S. DEPARTMENT OF COMMERCE
Clearinghouse for Federal Scientific and Technical Information
Springfield, Va. 22151

IX/16/3

TABLE OF CONTENTS

Preface viii

Part One
ABSOLUTE STELLAR MAGNITUDES AND STELLAR MASSES

Chapter I. ABSOLUTE STELLAR MAGNITUDES (O.A. Mel'nikov) 1
§ 1. Introduction. Brightness and color of stars 1
§ 2. The bolometric correction 7
§ 3. Average, effective, and isophotic wavelengths and their calculation 8
§ 4. Modern color systems. Interstellar absorption 16
§ 5. Radiation fluxes (illuminances) and stellar brightness 20
§ 6. Methods of practical determination of absolute stellar magnitudes 25
§ 7. Spectral classification of cold stars. Criteria of the MK classification 40

Chapter II. STELLAR MASSES (O.A. Mel'nikov) 44
§ 8. General 44
§ 9. Determination of mass from the red shift 47
§ 10. Mass—luminosity relation 49

Part Two
BINARY STARS

Chapter III. VISUAL BINARIES (A.N. Deich) 56
§ 11. A historical survey 56
§ 12. Observations of visual binaries 59
§ 13. Determination of orbits of visual binaries 62
§ 14. Determination of masses of binaries 69
§ 15. Dark companions 71
§ 16. Wide pairs 74
§ 17. Triple and multiple stars 77
§ 18. The statistics of double and multiple stars 79

Chapter IV. SPECTROSCOPIC BINARIES (V.A. Krat) 84
§ 19. Radial velocity curves 84
§ 20. Fundamental equations for the determination of orbits. The Lehmann—Filhés method 86
§ 21. The methods of Irwin, Schwarzschild, and others 90
§ 22. Improving the elements by the least squares method 91

Chapter V. PHOTOMETRIC BINARIES (ECLIPSING VARIABLES) (V.A. Krat) 93
§ 23. The light curve 93
§ 24. Models of binaries 94
§ 25. The model of spherical stars. The U hypothesis 96
§ 26. The model of spherical stars. The D hypothesis 101
§ 27. Intermediate cases of limb darkening 103
§ 28. Modifications of Russell's method. The method of Fetlaar and the Kopal—Piotrovski method 105

§ 29. Determination of corrections to relative elements by the least squares method 108
§ 30. A direct method for estimating the dispersions of relative elements 112
§ 31. Elliptical orbits (the model of spherical stars) 112
§ 32. The motion of the line of apsides. The presence of a third body 117
§ 33. Tidal deformations (the model of ellipsoidal stars) 118
§ 34. The sphere—ellipsoid model 122
§ 35. The phase effect 124
§ 36. Non-stable stars 127
§ 37. Classification of close pairs 129
§ 38. Some remarks on stellar evolution in close binaries 133

Part Three
VARIABLES AND NOVAE

Chapter VI. INTRODUCTION (V.G. Gorbatskii) 139
§ 39. Methods of study of variables 140
§ 40. Designation and catalogues 142
§ 41. Classification of variables 143

Chapter VII. CEPHEIDS (O.A. Mel'nikov) 148
§ 42. Long-period cepheids 148
§ 43. RR Lyrae stars 173
§ 44. W Virginis stars 182
§ 45. β Canis Majoris stars 184
§ 46. General characteristics of the cepheid population 188

Chapter VIII. LONG-PERIOD, SEMIREGULAR, AND IRREGULAR HIGH-LUMINOSITY VARIABLES (V.G. Gorbatskii) 219
§ 47. Long-period variables 219
§ 48. Semiregular variables—RV Tauri stars 227
§ 49. Semiregular and irregular variable stars of other types 230
§ 50. Spectroscopic variables 232

Chapter IX. LOW- AND MEDIUM-LUMINOSITY VARIABLES (V.G. Gorbatskii) 238
§ 51. Z Andromedae stars 238
§ 52. U Geminorum and Z Camelopardalis stars 241
§ 53. R Coronae Borealis stars 244
§ 54. RW Aurigae and T Tauri stars 246
§ 55. UV Ceti stars 249

Chapter X. STARS WITH VARIABLE SPECTRA (B.A. Vorontsov-Vel'yaminov) 251
§ 56. Spectroscopic variables 251
§ 57. Stars with variable bright lines 251
§ 58. Variable magnetic fields in stars 254

Chapter XI. NOVAE, SUPERNOVAE AND NOVA-LIKE STARS (B.A. Vorontsov-Vel'yaminov) 257
§ 59. Novae and nova-like stars 257
§ 60. Recurrent novae 258
§ 61. Light curves of novae 260
§ 62. Main spectral stages 262
§ 63. Nebulae illuminated and ejected by novae 266
§ 64. Physical conditions in nebulae 269
§ 65. Physical effects in the eruption of novae 270
§ 66. Novae in other galaxies and their position on the Hertzsprung—Russell diagram. Spatial distribution 275
§ 67. Nova-like stars 278

§ 68. P Cygni stars 279
§ 69. Late-type nova-like stars 282
§ 70. The mechanism of nova eruption 284
§ 71. Supernovae 285

Part Four
DIFFUSE MATTER

Chapter XII. GASEOUS NEBULAE (B. A. Vorontsov-Vel'yaminov) 293
§ 72. Description of planetary nebulae 293
§ 73. The spectra and the light of planetary nebulae 295
§ 74. Distances, sizes, luminosities, and spatial structure of planetary nebulae 298
§ 75. The nature of the nuclei 300
§ 76. Emission nebulae and hydrogen fields 302
§ 77. Interstellar gas 310
§ 78. H I and H II regions 315
§ 79. Radio observations of the gaseous diffuse matter 317

Chapter XIII. DUST NEBULAE (B. A. Vorontsov-Vel'yaminov) 320
§ 80. Interstellar dust and dark nebulae 320
§ 81. Bright dust nebulae 328

Part Five
THEORY OF STELLAR ATMOSPHERES AND GASEOUS NEBULAE

Chapter XIV. STELLAR PHOTOSPHERES (V. V. Sobolev) 332
§ 82. Radiative equilibrium of a stellar photosphere 332
§ 83. Brightness distribution over the stellar disk 339
§ 84. Energy distribution in the continuous spectrum 341
§ 85. Theory of stellar photospheres with a frequency-dependent absorption coefficient 345
§ 86. The structure of stellar photospheres 352

Chapter XV. STELLAR ATMOSPHERES (V. V. Sobolev) 357
§ 87. The Schwarzschild—Schuster model 357
§ 88. The equivalent line width as a function of the number of absorbing atoms 361
§ 89. Chemical composition of stellar atmospheres 366
§ 90. Line broadening 369
§ 91. Eddington's model 374
§ 92. Central intensities of absorption lines 378
§ 93. Absorption lines with incoherent scattering 383
§ 94. Interpretation of the spectral sequence 387

Chapter XVI. GASEOUS NEBULAE (V. V. Sobolev) 392
§ 95. The origin of nebular light 392
§ 96. Determination of stellar temperatures 396
§ 97. Ionization in nebulae 400
§ 98. Emission line intensities 404
§ 99. Forbidden lines in nebular spectra 407
§ 100. Nebular temperatures 411
§ 101. Masses, densities, and chemical composition 413
§ 102. Radiation pressure in nebulae 415
§ 103. The continuous nebular spectrum 419

Chapter XVII. STARS WITH BRIGHT SPECTRAL LINES (V. V. Sobolev) 425
§ 104. Early-type stars with bright lines 425
§ 105. Emission line profiles 427

§ 106. Emission line intensities 430
§ 107. Be stars 433
§ 108. Wolf—Rayet stars 435
§ 109. Late-type stars with bright lines 439

Part Six
STELLAR SYSTEMS

Chapter XVIII. STELLAR STATISTICS. GALACTIC STRUCTURE (V. V. Sobolev) 443
§ 110. Galactic concentration 443
§ 111. Galactic coordinates 443
§ 112. Luminosities and absolute stellar magnitudes. Spectrum—luminosity diagram 445
§ 113. Color—luminosity diagram 448
§ 114. Mass—luminosity diagram 451
§ 115. Luminosity function 453
§ 116. Apparent magnitude function 455
§ 117. The total number of stars in the Galaxy 459
§ 118. Star density function. Determination of star density near the Sun 460
§ 119. Fundamental equations of stellar statistics 465
§ 120. General solution of the first equation of stellar statistics 466
§ 121. Approximate solution of the first equation of stellar statistics 467
§ 122. Light absorption 468
§ 123. Selective absorption 469
§ 124. Interstellar absorption and relation of M to m 471
§ 125. Interstellar absorption and the fundamental equation of stellar statistics 472
§ 126. Mean absorption function 472
§ 127. The Vashakidze—Oort method for the determination of star density 474
§ 128. The significance of dark nebulae in interstellar absorption 475
§ 129. Theory of field brightness fluctuations 480
§ 130. Surface brightness fluctuations of the Galaxy 484
§ 131. Main features of galactic structure 486
§ 132. Two types of stellar populations. Subsystems of the Galaxy 489
§ 133. The spiral structure of the Galaxy 491
§ 134. Radio structure of the Galaxy 492
§ 135. The Local Group 496

Chapter XIX. STELLAR MOTIONS (A. N. Deich and T. A. Agekyan) 499
§ 136. General considerations 499
§ 137. The observed spatial velocity components and their relation to x, y, z 500
§ 138. Secular variation of proper motion, parallax, and radial velocity of a star due to its motion in space 503
§ 139. Statistical relations between the mean spatial velocity of a star and its radial and tangential components 505
§ 140. Proper motion components v and τ 506
§ 141. Average secular and average annual parallaxes of stars 508
§ 142. Determination of the Sun's apex from proper motions 511
§ 143. Determination of the elements of the Sun's motion from radial velocities 513
§ 144. The K effect 514
§ 145. Kapteyn's "selected areas" 515
§ 146. The centroid. Residual star velocities 517
§ 147. Distribution of residual velocities 519
§ 148. Asymmetry of stellar motions 525
§ 149. The effect of galactic rotation on radial velocities and proper motions 529
§ 150. The differential field of centroid velocities. Its effect on the observed stellar motions . 531
§ 151. The centroid field for pure galactic rotation 537
§ 152. Galactic rotation from radio observations 538
§ 153. The spin of the stars 544

Chapter XX. STELLAR DYNAMICS (T.A. Agekyan) 549
§ 154. Basic premises 549
§ 155. The fundamental equation of stellar dynamics for the regular field 552
§ 156. The solution of the fundamental equation of stellar dynamics using a known potential 555
§ 157. The hydrodynamic equations for stellar systems 558
§ 158. The solution of the fundamental equation of stellar dynamics using a known velocity distribution function 560
§ 159. Variation of the residual velocity of a star. Circular and almost circular orbits 565
§ 160. The virial theorem and its consequences 571
§ 161. Irregular forces. Variation of the absolute velocities of stars following binary encounters 574
§ 162. The probability of a binary encounter with a given change in absolute velocity 577
§ 163. The general problem of velocity variation in the irregular field. Relaxation time 583
§ 164. Distribution function of the residual velocities and the rate of dissipation of stellar systems 585
§ 165. The most probable distribution in phase space 587
§ 166. Dissipation of stellar systems 590
§ 167. Evolution of dissipating stellar systems 594
§ 168. General features in the evolution of stellar systems 596

Chapter XXI. STAR CLUSTERS (T.A. Agekyan) 601
§ 169. Introduction 601
§ 170. Open star clusters 601
§ 171. Moving clusters 610
§ 172. Stellar associations 612
§ 173. Globular clusters 615
§ 174. Stability of star clusters 625

Chapter XXII. EXTRAGALACTIC ASTRONOMY (T.A. Agekyan) 631
§ 175. Introduction 631
§ 176. Hubble's classification of galaxies 631
§ 177. Distances of galaxies 638
§ 178. Apparent magnitudes, absolute magnitudes, diameters, and surface brightness of galaxies 640
§ 179. Spectra, color indices, and stellar populations of galaxies. Morgan's classification ... 647
§ 180. The frequency of galactic types 651
§ 181. Distribution of the flattening of galaxies 652
§ 182. The red shift in galactic spectra 655
§ 183. Revision of the scale of extragalactic distances 659
§ 184. Rotation of galaxies 663
§ 185. Determination of galactic masses from rotation 667
§ 186. The Magellanic Clouds and the Andromeda Nebula 671
§ 187. Apparent distribution of galaxies 678
§ 188. Double and multiple galaxies 683
§ 189. Masses of double galaxies 686
§ 190. Clusters of galaxies 687
§ 191. Determination of galactic masses in clusters. The mass—luminosity relation in star systems 693
§ 192. Radio emission from galaxies 696
§ 193. The Metagalaxy. The structure of the Universe 701

PREFACE

The second edition of the first volume of this Course in Astrophysics and Stellar Astronomy was published in 1951. Written mainly by the staff of the Pulkovo Observatory, it dealt with methods and instruments used in astrophysical work. According to the original plan, the second volume was to follow soon after that, presenting the results of planetary and stellar observations, their interpretation, and a discussion of the structure of star systems. The Editorial Board, however, met with certain difficulties in their efforts to find authors for the different chapters. The material moreover proved to be too extensive for one volume, and it therefore had to be divided into two volumes, one dealing with the star universe, and the other with the planetary system. All this led to a considerable delay in publication, and it was 1962 before the second volume of the Course, dealing with stars and stellar systems, could be delivered to the printers. Although logically the second volume of the Course should have been devoted to the solar system, it was our intention to avoid further delay in fixing the present publication schedule. The change in the natural order of volumes may yet prove of some advantage, since our knowledge of the planetary system, the Moon in particular, is rapidly expanding and it is hoped that the latest findings of planetary research will be incorporated in Volume III.*

* [Volume III of this Course was published in Russian in 1964 and translated into English under the title Physics of the Solar System in 1966 by Israel Program for Scientific Translations, NASA TT F-342, TT 66-51022, IPST Cat. No. 1526.]

Part One

ABSOLUTE STELLAR MAGNITUDES AND STELLAR MASSES

O.A. MEL'NIKOV

Chapter I

ABSOLUTE STELLAR MAGNITUDES

§1. **Introduction. Brightness and color of stars.** Stars appear as bright points to the eye. The absolute brightness (or absolute luminance) of a star (E) is defined as the illuminance produced by the star on unit surface perpendicular to the light rays at a standard distance of 10 parsecs (see Vol. I, pp. 346—349). In accordance with the relation $M = -2.5 \lg E + \text{const}$, the absolute magnitude (M) of a star is defined as the apparent magnitude at a standard distance of 10 parsecs (in the absence of light absorption), which corresponds to a parallax of $0''.1$.

Some very bright early-type B stars in the Galaxy and the large Magellanic Cloud are known to have absolute magnitudes reaching $-9^m.5$. At the other extreme of the brightness scale, we find late-type stars with absolute magnitudes of up to $+19^m.0$. Since $M_{v,\odot} = 4^m.73$, the stars at the two extreme ends of the scale are respectively 500,000 brighter and dimmer than the Sun (in terms of absolute brightness or luminosity, i.e., the total radiation from the entire surface). The Sun generates radiant energy at a power of 2 erg/sec per 1 g of its mass. The brightest stars, on the other hand, radiate thousands of times as much, whereas the radiation of the dimmest stars is three orders of magnitude less powerful than that of the Sun; note that the mass differences are only one order of magnitude. Hence it follows that the life expectancy of the bright stars is relatively small ($10^5 - 10^8$ years), and that of the faint stars large ($10^9 - 10^{12}$ years).

We gave the extreme limits of the absolute stellar magnitudes. The commonly observed stars have absolute visual magnitudes of approximately between -6^m and $+10^m$ or even $+15^m$ (white dwarfs). The difference in stellar magnitudes equal to $-6^m - 15^m = -21^m$ corresponds to a luminosity ratio of 250,000,000:1!

The so-called supernovae — a special type of erupting non-stationary stars — may reach even higher luminosities, though only at the maximum. For example, the supernova which erupted in 1955 in the spiral galaxy NGC 23 had an absolute visual magnitude of $-18^m.0$ and was only a few stellar magnitudes weaker than the integrated absolute magnitude of the entire galaxy, which probably contains billions of average Sun-type stars.

Direct observations only give the apparent stellar magnitude m. The absolute magnitude of a star can be calculated from the apparent magnitude if we know the distance of the star r or its parallax p.* Unfortunately, the apparent magnitude is further reduced by interstellar

* The absolute magnitude can also be found approximately from the observed spectrum (see below).

absorption (so that the algebraic value of m comes out too high). In the following we assume for simplicity that the appropriate absorption corrections have been introduced into the relevant figures. Let a star of radius R emit in the spectral region of wavelength λ (cm) an energy flux $I_\lambda(\theta)$ per unit surface area (cm^2) in unit solid angle (sterad) in unit time (sec) in unit spectral interval ($\Delta\lambda = 1$ cm). This energy flux is called the specific spectral intensity or simply the intensity in the direction θ to the normal (the radius); its dimensions are clearly erg/sec · cm^2 · · sterad. The total radiation emitted from unit surface area into the observer's hemisphere (the radiation flux) is given (for $\Delta\lambda = 1$ cm) by

$$H_\lambda = \int I_\lambda \cos\theta \, d\omega = \int_{\theta=0}^{\frac{\pi}{2}} \int_{\varphi=0}^{2\pi} I_\lambda(\theta) \cos\theta \sin\theta \, d\theta \, d\varphi \text{ erg/sec} \cdot \text{cm}^2. \qquad (1)$$

Let $\bar{I}_\lambda$ be the intensity averaged over the star's disk:

$$\bar{I}_\lambda = \frac{\int_0^{\frac{\pi}{2}} \int_0^{2\pi} I_\lambda(\theta) \cos\theta \sin\theta \, d\theta \, d\varphi}{\int_0^{\frac{\pi}{2}} \int_0^{2\pi} \cos\theta \sin\theta \, d\theta \, d\varphi} = 2 \int_0^{\frac{\pi}{2}} I_\lambda(\theta) \cos\theta \sin\theta \, d\theta. \qquad (2)$$

Since the projection of a surface element of the star onto the disk is $R^2 \cos\theta \sin\theta \, d\theta \, d\varphi$, we have

$$\pi R^2 \bar{I}_\lambda = \int_0^{\frac{\pi}{2}} \int_0^{2\pi} I_\lambda(\theta) R^2 \cos\theta \sin\theta \, d\theta \, d\varphi. \qquad (3)$$

Comparison of (1) and (3) gives for the radiation flux in unit spectral interval ($\Delta\lambda = 1$ cm)

$$H_\lambda = \pi \bar{I}_\lambda \text{ erg/sec} \cdot \text{cm}^2 \text{ (or W/cm}^2\text{)}. \qquad (4)$$

The radiation flux in a spectral interval from λ to $\lambda + d\lambda$, i.e., a wavelength interval of $d\lambda$, is naturally equal to

$$H_\lambda \, d\lambda = \pi \bar{I}_\lambda \, d\lambda \text{ erg/sec} \cdot \text{cm}^2 \text{ (or W/cm}^2\text{)}. \qquad (5)$$

The monochromatic energy flux radiated every second by the entire surface of the star in all directions in a unit wavelength interval $\Delta\lambda = 1$ cm is $L_\lambda = 4\pi R^2 H_\lambda$ erg/sec (or watt). The energy flux in a spectral interval $d\lambda$ is

$$L_\lambda \, d\lambda = 4\pi R^2 \cdot H_\lambda \, d\lambda = 4\pi R^2 \cdot \pi \bar{I}_\lambda \, d\lambda \text{ erg/sec (or W)}. \qquad (6)$$

L_λ is the monochromatic luminosity of the star in a spectral interval $\Delta\lambda = 1$ cm. In practice, we are of course concerned not with the theoretical luminosity L_λ but with the luminosity L_{λ_1, λ_2} which is defined for

a blackbody source by an integral of the form

$$L_{\lambda_1,\lambda_2} = 4\pi R^2 \int_{\lambda_1}^{\lambda_2} J_\lambda(T)\, d\lambda = 4\pi R^2 \int_{\lambda_1}^{\lambda_2} \frac{\pi C_1}{\lambda^5} \frac{d\lambda}{e^{c_2/\lambda T}-1} \approx$$
$$\approx 4\pi R^2 [\pi C_1 (e^{c_2/\lambda_0 T}-1)^{-1} \cdot \lambda_0^{-5} (\lambda_2 - \lambda_1)]. \quad (7)$$

In calculating L_λ we took $\Delta\lambda = 1$ cm. In visual and photographic observations $\lambda_2 - \lambda_1$ is approximately equal to $3 \cdot 10^{-5}$ cm and $2.05 \cdot 10^{-5}$ cm, respectively, i.e., relatively small, and we may thus take $J_\lambda(T)$ to be constant in this interval, corresponding to its value for some λ_0 near the midpoint of the interval $(\lambda_2 - \lambda_1)$. The integration limits, or the interval $(\lambda_2 - \lambda_1)$, determine the photometric system of luminosities. For $\lambda_1 \approx 4000$ Å, $\lambda_2 \approx 7000$ Å, we obtain the visual L_v; for $\lambda_1 = 2950$ Å, $\lambda_2 = 5000$ Å the photographic L_{ph}, and for $\lambda_1 = 0$, $\lambda_2 = \infty$ the bolometric luminosity L_b. The absolute stellar magnitudes are correspondingly defined as $M = -2.5 \lg L +$ + const. In practice the situation is further complicated by the fact that the detector response, even for a blackbody source, is not proportional to $L \sim \int_{\lambda_1}^{\lambda_2} J_\lambda(T)\, d\lambda$ but actually to $\sim \int_0^\infty J_\lambda(T)\, s(\lambda)\, d\lambda$, where the function $s(\lambda)$ is a product of at least three unknown functions (see below). At this stage, however, we will use theoretical luminosities only. The function

$$L \equiv L_b = \int_0^\infty L_\lambda d\lambda \text{ erg/sec (or W)} \quad (8)$$

is called the integrated (over the spectrum), total, or else bolometric luminosity of a star. For the Sun, e.g.,

$$L_{b,\odot} = 3.840 \cdot 10^{33} \text{ erg/sec} = 3.840 \cdot 10^{26} \text{ W}.$$

The illuminance produced by a star distant r pc (parsecs) from the Earth on a surface element perpendicular to the energy flux just outside the Earth's atmosphere is (for $\Delta\lambda = 1$ cm)

$$e_\lambda = \frac{L_\lambda}{4\pi r^2} = \frac{R^2}{r^2} H_\lambda = \frac{R^2}{r^2} \pi \bar{I}_\lambda \text{ erg/sec} \cdot \text{pc}^2 \text{ or W/pc}^2. \quad (9)$$

This quantity is otherwise known as the surface brightness. If the same star is assumed to be at a standard distance $r = 10$ pc, the absolute surface brightness in the same units is given by

$$E_\lambda = \frac{L_\lambda}{4\pi (10)^2} = \frac{R^2 (nc^2)}{10^2} H_\lambda = \left(\frac{R}{10}\right)^2 \pi \bar{I}_\lambda. \quad (10)$$

e_λ and E_λ were computed for a unit wavelength interval $\Delta\lambda = 1$ cm. Their ratio can be written in the form

$$\frac{E_\lambda}{e_\lambda} = \left(\frac{r}{10}\right)^2 = \left(\frac{0''.1}{p}\right)^2. \quad (11)$$

Here p is the stellar parallax in seconds of arc, corresponding to a distance r pc, and $0''.1$ is the parallax corresponding to a distance of 10 pc. Let the

illuminance E_λ correspond to absolute stellar magnitude M_λ, and e_λ to the apparent magnitude m_λ. Then from Pogson's formula we get

$$\frac{E_\lambda}{e_\lambda}=\left(\frac{1}{2.512}\right)^{M_\lambda-m_\lambda}=10^{-0.4(M_\lambda-m_\lambda)}. \tag{12}$$

Comparing (11) and (12) and taking the logarithm, we obtain

$$\lg\frac{E_\lambda}{e_\lambda}=2\lg r-2=-2-2\lg p=-0.4(M_\lambda-m_\lambda). \tag{13}$$

Hence

$$M_\lambda=m_\lambda+5-5\lg r=m_\lambda+5+5\lg p. \tag{14}$$

This is a well known relation often used in stellar astronomy. It relates the absolute and the apparent stellar magnitudes through the distance or the parallax (see Vol. I, p. 349).

By (6) and (10), we write for the absolute stellar magnitude

$$M_\lambda=\text{const}-2.5\lg E_\lambda=\text{const}-5\lg R-2.5\lg H_\lambda=\text{const}-5\lg R-2.5\lg\pi\cdot\bar{I}_\lambda. \tag{15}$$

For two wavelengths λ_1 and λ_2 we have

$$\left.\begin{aligned}M_{\lambda_1}&=c_{\lambda_1}-5\lg R-2.5\lg\pi\cdot\bar{I}_{\lambda_1},\\ M_{\lambda_2}&=c_{\lambda_2}-5\lg R-2.5\lg\pi\cdot\bar{I}_{\lambda_2}.\end{aligned}\right\} \tag{16}$$

Here c_{λ_1} and c_{λ_2} are the appropriate constants.

We form the difference

$$M_{\lambda_1}-M_{\lambda_2}=-2.5\lg\pi(\bar{I}_{\lambda_1}-\bar{I}_{\lambda_2})+(c_{\lambda_1}-c_{\lambda_2}). \tag{17}$$

This difference is called the color index or simply the color of the star. The wavelengths λ_1 and λ_2 determine, in particular, the system of stellar magnitudes, colors, etc.

If a star emits as a blackbody of temperature T, we have $J_{\lambda_i}(T)=\bar{I}_{\lambda_i}=I_{\lambda_i}$. According to Planck's equation, the blackbody radiation from a unit surface area* in a wavelength interval $\Delta\lambda=1$ cm is

$$J_{\lambda_i}(T)=\frac{2hc^2}{\lambda_i^5}\frac{1}{e^{\frac{hc}{\lambda_i kT}}-1}=\frac{C_1}{\lambda_i^5}\frac{1}{e^{\frac{c_2}{\lambda_i T}}-1}=\frac{C_1}{\lambda_i^5}e^{-\frac{c_2}{\lambda_i T}}(1-e^{-\frac{c_2}{\lambda_i T}})^{-1}\ \text{erg/sec}\cdot\text{cm}^2\cdot\text{sterad}. \tag{18}$$

The constants C_1 and c_2 have the following values:

$$C_1=2hc^2=1.191\cdot10^{-5}\ \text{erg}\cdot\text{cm}^2/\text{sec}\cdot\text{sterad}=1.191\cdot10^{-12}\ \text{W}\cdot\text{cm}^2/\text{sterad}$$

and

$$c_2=\frac{hc}{k}=1.4385\ \text{cm}\cdot\text{deg},$$

the dimensions of $J_{\lambda_i}(T)$ are the same as of $\bar{I}_\lambda$ (erg/sec · cm^2 or W/cm^2), since in our case they were computed for a unit wavelength interval $\Delta\lambda=1$ cm.

* $\pi J_{\lambda_i}(T)$ is called the blackbody flux density, the blackbody radiation intensity or the blackbody luminosity (see Vol. I, p. 451).

Inserting (18) in (15), we get

$$M_\lambda = c_\lambda - 5\lg R + 2.5\lg e\cdot\frac{c_2}{\lambda T} - 2.5\lg\left(1-e^{-\frac{c_2}{\lambda T}}\right)^{-1} =$$

$$= c_\lambda - 5\lg R + 1.086\cdot\frac{c_2}{\lambda T} + x_\lambda = c_\lambda - 5\lg R + \frac{1.5600}{\lambda T} + x_\lambda. \quad (19)$$

The last term in the right hand side (x_λ) is zero if Wien's approximation is applicable to Planck's equation, i.e., if $e^{-\frac{c_2}{\lambda T}} \ll 1$, or alternatively $T > 3000°$, $\lambda > 3000$ Å (for $\frac{c_2}{\lambda T} \ll 1$ the Rayleigh—Jeans law is applicable to the radiation). In general, however, x_λ is never large. For $\frac{1.5600}{\lambda T} = 5, 4, 3, 2$, and 1, $x_\lambda =$ $= 0.01$, 0.03, 0.07, 0.19, and 0.55. For $\lambda = 5000$ Å and $T\cdot 10^{-3} = 4, 6, 8,$ 10, 15, 20, $\left(1-e^{-\frac{c_2}{\lambda T}}\right)^{-1}$ takes the values 1.001, 1.009, 1.029, 1.061, 1.174, and 1.315. In the Harvard photometric system, $\lambda_1 \equiv \lambda_{pg} \equiv \lambda_p = 4250$ Å and $\lambda_2 \equiv \lambda_{vis} \equiv \lambda_v = 5290$ Å for photographic and visual radiation, respectively. Using these values, we obtain from (19)

$$\left.\begin{aligned} M_p &= c_p - 5\lg R + \frac{36\,700}{T} + x_p, \\ M_v &= c_v - 5\lg R + \frac{29\,500}{T} + x_v. \end{aligned}\right\} \quad (20)$$

Hence for the color index

$$CI \equiv C = M_p - M_v = m_p - m_v = c_p - c_v + \frac{7200}{T} + (x_p - x_v). \quad (21)$$

The commonly used color system* has $C = 0$, i.e., $M_p = M_v$ for the A0 stars with color temperature T (A0) = 15,000°K. Therefore in this system $(c_p - c_v) = -\frac{7200}{15\,000} - (x_p - x_v) = -0.48 - 0.07 = -0.55$. Thus

$$C = M_p - M_v = m_p - m_v = \frac{7200}{T} - 0.55. \quad (22)$$

The color C varies from $-0^m.3$ to $+2^m.0$ for the hottest and the coldest stars, respectively.

Since the energy distribution in a star's spectrum, and hence the magnitudes M_p, M_v, are determined by the color temperature, and not the effective temperature, we use in (22) $T = T_c$, and not $T = T_e$.

The color temperature is 15,000° and higher for the white-blue stars and 3000° or less for the dark-red stars. Since $c_v = c_p + 0^m.55$, we may write

$$\left.\begin{aligned} M_p &= M_0 - 5\lg\frac{R}{R_\odot} + \frac{36\,700}{T} + x_p, \\ M_v &= M_0 + 0.55 - 5\lg\frac{R}{R_\odot} + \frac{29\,500}{T} + x_v. \end{aligned}\right\} \quad (23)$$

* In photometric systems the zero point of photographic and visual magnitudes is selected as follows: within the range of apparent visual stellar magnitudes $m_v = 5^m.5 - 6^m.5$, the mean Harvard photographic magnitudes of stars of Harvard spectral class A0 coincide with the mean Harvard visual magnitudes of stars of the same class A0, i.e., the color index is $C(A0) = \overline{m}_{pg} - \overline{m}_v = 0$. This technique for zero point determination, however, is now considered inadequate.

The common constant M_0 can be determined in reference to the Sun. For the Sun $m_v = -26^m.84$ (G. Kuiper) and the mean distance

$$r = 1.495 \cdot 10^{13} \text{ cm} = \frac{1.495 \cdot 10^{13}}{3.086 \cdot 10^{18}} \text{pc} = 4.844 \cdot 10^{-6} \text{pc}.$$

Hence

$$M_v(\odot) \equiv M_{v,\odot} = m_v(\odot) + 5 - 5 \lg r_\odot(\text{pc}) = -26.84 + 5 + 26.57 = +4^m.73. \quad (24)$$

By (23), taking $T \approx 6{,}000°\text{K}$, we thus obtain

$$M_v = 4^m.73 = M_0 + 0.55 + 4.92 - 0.01 = M_0 + 5.46. \quad (25)$$

Hence, $M_0 = -0^m.73$ and $(M_0 + 0.55) = -0^m.18$. Therefore (23) is rewritten in the form

$$\left.\begin{aligned} M_p &= -0.73 - 5 \lg \frac{R}{R_\odot} + \frac{36\,700}{T} + x_p, \\ M_v &= -0.18 - 5 \lg \frac{R}{R_\odot} + \frac{29\,500}{T} + x_v. \end{aligned}\right\} \quad (26)$$

In our preceding calculations we used Kuiper's magnitude $m_v(\odot) = -26^m.84$ obtained in 1938. Many new determinations have been published since then. The most exact determination, followed by a comprehensive discussion, was carried out by Martynov in 1959. In the color V of the U, B, V system he obtained for the isophotic line $\lambda = 0.55\mu$, $m_v(\odot) = -26^m.80$, i.e., virtually the same magnitude as that used in our calculations. Martynov also determined the stellar magnitude equivalent of 1 meter-candle (international lux) outside the atmosphere, which was found to be $-13^m.78$.

From the above equations, we can derive an approximate estimate for the stellar diameter (radius). Indeed, from the definition of the parallax p'' and the angular diameter D'' we have

$$D'' = p'' \frac{D}{A} = p'' \frac{2R}{A} = p'' \frac{2\frac{R}{R_\odot}}{\frac{A}{R_\odot}} = p'' \frac{\frac{R}{R_\odot}}{\frac{A}{2R_\odot}} = p'' \frac{\frac{R}{R_\odot}}{107}. \quad (27)$$

Here $\frac{A}{2R_\odot} = 107$ (1 astronomical unit expressed in solar diameters). Hence,

$$\lg D'' = \lg p'' + \lg \frac{R}{R_\odot} - 2.029. \quad (28)$$

Using (14) and (26) for M_v, we find

$$\lg p \equiv \lg p'' = -0.200 m_v - \lg \frac{R}{R_\odot} + \frac{5900}{T} - 1.036 + 0.200 x_v. \quad (29)$$

Finally, by (28) and (29) we obtain the approximate equality

$$\lg D'' = -0.200 m_v + \frac{5900}{T} - 3.070 + 0.200 x_v. \quad (30)$$

This relation is used in approximate calculation of stellar diameters and radii.

§2. The bolometric correction. The so-called bolometric correction is a quantity of considerable significance in astrophysics. It is equal to the difference between the visual M_v and the bolometric M_b absolute (or apparent) stellar magnitudes:

$$\Delta M_b = M_v - M_b = \Delta m_b = m_v - m_b = -B.C. \tag{31}$$

Initially this correction was computed for a blackbody and the zero point was defined as the minimum of the $\Delta M_v(T)$ curve around 6800°K, where zero bolometric correction was assumed. The integrated-light flux emitted into the observer's hemisphere from unit surface area of a blackbody source is

$$H = \pi \bar{I} = \pi \int_0^\infty \bar{I}_\lambda \, d\lambda = \sigma T^4 \text{ (erg/sec} \cdot \text{cm}^2 \text{ or W/cm}^2\text{)}. \tag{32}$$

The integrated flux from the entire surface of the star — its bolometric luminosity — is therefore given by

$$L_b = 4\pi R^2 H = 4\pi R^2 \pi \bar{I} = 4\pi R^2 \sigma T^4 \text{ (erg/sec or W)}. \tag{33}$$

The absolute bolometric stellar magnitude (using the zero minimum of $M_v - M_b$ at $T \approx 6800°$K) is equal to

$$M_b = c_b - 5 \lg \frac{R}{R_\odot} - 10 \lg T = M_0 + 43.18 - 5 \lg \frac{R}{R_\odot} - 10 \lg T. \tag{34}$$

By (23) and (34)

$$\Delta M_b = M_v - M_b = M_0 + 0^m.55 - c_b + \frac{29\,500}{T} + 10 \lg T + x_v. \tag{35}$$

Since in the minimum at $T \approx 6800°$K, we have $\Delta M_b \equiv 0$, the constant can be found. Finally we obtain for the bolometric correction (sometimes designated $B.C. = -\Delta M_b$)

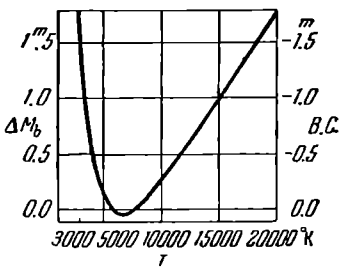
$$-B.C. = \Delta M_b = M_v - M_b = m_v - m_b = -42.63 + 10 \lg T + \frac{29\,500}{T} + x_v. \tag{36}$$

The bolometric correction enables us to determine the bolometric (integrated) magnitude of a star from its visual magnitude. This is particularly important as the bolometric magnitude cannot be determined directly from observations. This is associated with the inaccessibility of the far infrared and of the wavelengths shorter than 0.295μ to observations (these wavelengths are absorbed in the Earth's atmosphere). The Sun is the only star for which these wavelengths are accessible, using special instruments from balloons, rockets, and artificial Earth satellites.

A plot of the bolometric correction is shown in Figure 1. For the Sun at $T_\odot \approx 5780°$K, for example, $\Delta M_b = +0^m.07$, and therefore $M_b = M_v - \Delta M_b = +4^m.73 - 0^m.07 = +4^m.66$. By (22), for $T_\odot \approx 5780°$K, the color $C_{p-v} = +0^m.69$ and therefore for the Sun $M_p = M_v + C_{p-v} = +5^m.42$. For a normal star, the Hertzsprung—Russell diagram (HR diagram) gives for $M_v = +4^m.60$ a color index $C_{p-v} = +0^m.65$, which corresponds to the dG0 spectrum.*

* According to the latest data the spectral class of the Sun in the MK system (see below) is G2V.

Note that the Fraunhofer spectrum of the Sun is substantially variable over the disk; it corresponds to G1, G4, G9p, and K0p for $\sin\theta = 0$, 0.75, 0.945, and 0.985, respectively (W. Morgan, Ph. Keenan, 1939).



For late-type stars the bolometric corrections may be approximately derived from observations, since the main (in terms of energy) region of the continuum is situated in a region which is accessible to bolometers, radiometers, and other thermal detectors. The situation becomes more complicated for early-type stars, where a substantial fraction of energy is radiated in the inaccessible ultraviolet. In this case the correction is computed from (36). This approach, however, is again unreliable, since the energy distribution in the continuum departs from the blackbody spectrum. More exact results are obtained in theoretical astrophysics by computing the energy distribution in the continuum for a certain model of the atmosphere at a given temperature.

We will return to this problem at a later stage, as it has direct bearing on the question of the scale of effective stellar temperatures. The main difficulty is that the observations give the magnitudes M_v and m_v which are not monochromatic. The rare exceptions are only those cases when the magnitudes are determined from continuous spectra obtained with prism cameras or slitless spectrographs.

§3. **Average, effective, and isophotic wavelengths and their calculation.** In practice we are generally dealing with heterochromatic, many-colored stellar magnitudes. This is so because the light detector responding to the incident luminous flux from the star has a response characteristic which theoretically extends from the shortest to the longest wavelengths; in practice the response band is finite, ranging between some wavelengths λ_1 and λ_2, with a peak near some intermediate wavelength. Moreover, the luminous flux of the star is attenuated 1) in the Earth's atmosphere, 2) by absorption and reflection losses in the telescope optics, 3) by absorption in light filters, etc. This problem was discussed in some detail in Vol. I of this Course. If $\mathfrak{D}$ is the diameter of the telescope objective, then by (9) the detected intensity, i.e., the radiation flux measured by the detector response, is

$$i \sim \frac{\pi\mathfrak{D}^2}{4}\int_0^\infty e_\lambda s(\lambda)\,d\lambda \sim \frac{\pi\mathfrak{D}^2}{4}\int_0^\infty \bar{I}_\lambda(T)\,s(\lambda)\,d\lambda \sim \int_0^\infty \bar{I}_\lambda(T)\,s(\lambda)\,d\lambda \approx \int_{\lambda_1}^{\lambda_2} \bar{I}_\lambda(T)\,s(\lambda)\,d\lambda. \quad (37)$$

Here $\bar{I}_\lambda$ is a function which describes the energy distribution in the star's continuum, and $s(\lambda)$ is a function which is a product of all the normalized functions describing the instrumental response and other factors (extinction of light in the atmosphere, in the telescope, etc.). Hence, $\frac{\pi\mathfrak{D}^2}{4}s(\lambda)$ is the fraction of the initial light flux penetrating through the Earth's atmosphere, the telescope optics, and the rest of the equipment to be picked up by the detector.

There is an obvious uncertainty in the measurement of i and thus in the determination of $M = -2.5 \lg i + \text{const}$ in the sense that we do not know to what spectral region they actually correspond (and do not know what wavelengths to insert in the above relations). We should choose some conditional wavelength so that the actual measured heterochromatic illuminance from the star, its apparent stellar magnitude, etc., coincide with the theoretically calculated monochromatic magnitudes at that wavelength. Approximate methods are generally applied to determine one of the following three wavelengths:

1) the weighted average wavelength of the system (for $I_\lambda(T) = \text{const}$, i.e., for uniform energy distribution in the spectrum)

$$\overline{\lambda} = \frac{\int_{\lambda_1}^{\lambda_2} \lambda I_\lambda(T)\, s(\lambda)\, d\lambda}{\int_{\lambda_1}^{\lambda_2} I_\lambda(T)\, s(\lambda)\, d\lambda} = \frac{\int_{\lambda_1}^{\lambda_2} \lambda\, s(\lambda)\, d\lambda}{\int_{\lambda_1}^{\lambda_2} s(\lambda)\, d\lambda}; \tag{38}$$

here, as in the following, the integration is mathematically carried out from 0 to ∞, whereas in practice the integration limits are λ_1 and λ_2, since outside the interval $(\lambda_2 - \lambda_1)$, $s(\lambda) = 0$;

2) the effective weighted average wavelength*

$$\bar{\lambda}_e = \lambda_e = \frac{\int_{\lambda_1}^{\lambda_2} \lambda \overline{I}_\lambda(T) s(\lambda)\, d\lambda}{\int_{\lambda_1}^{\lambda_2} \overline{I}_\lambda(T)\, s(\lambda)\, d\lambda}; \tag{39}$$

3) the isophotic wavelength λ_i. The latter is defined in different ways (see, e.g., Woolley and Stibbs, 1953). In particular, from the equation for blackbody radiation $J_{\lambda_i}(T)$ of color temperature $T \equiv T_c$,

$$J_{\lambda_i}(T_c) = \frac{1}{k} \int_{\lambda_1}^{\lambda_2} J_\lambda(T_c)\, s(\lambda)\, d\lambda = \text{const} \times \int_{\lambda_1}^{\lambda_2} J_\lambda(T_c)\, s(\lambda)\, d\lambda. \tag{40}$$

Here $\frac{1}{k}$ is a constant; taking numerical values for this constant, we obtain a family of λ_i values.

To avoid ambiguity the constant $\frac{1}{k}$ can be defined (Seares, Joyner 1943) by the equation

$$J_{\lambda_i}(T_c) = J_{\overline{\lambda}_E}(T_0) \frac{\int_0^\infty J_\lambda(T_c)\, s(\lambda)\, d\lambda}{\int_0^\infty J_\lambda(T_0)\, s(\lambda)\, d\lambda}. \tag{41}$$

$J_\lambda(T_0) = [J_\lambda(T_0)]_{\max}$, i.e., this function has a maximum at the weighted average or effective wavelength $\overline{\lambda}_E \equiv \lambda_{\max} \equiv \lambda_m$ corresponding to the temperature

$$T_0 = \frac{0.2897}{\overline{\lambda}_E} \equiv \frac{0.2897}{\lambda_{\max}} \equiv \frac{0.2897}{\lambda_m}.$$

* Note that in the literature the wavelength λ is often denoted by λ_e, and called the effective wavelength.

Clearly, for $T \equiv T_c = T_0$ we find $\lambda_i = \bar{\lambda}_E \equiv \lambda_{max} \equiv \lambda_m$. $\bar{\lambda}_E$ is generally so chosen that it corresponds to the T_c for A0 stars. The isophotic wavelengths are of great significance in astrophysics. Already their name indicates that the averaging of brightnesses — stellar magnitudes — over wavelengths, using the function $s(\lambda)$, gives a heterochromatic m which coincides with the calculated monochromatic m_{λ_i}. The concept of λ_i was introduced by A. Brill in 1929. In the particular case when the energy distribution in the continuum is a linear function in the relevant wavelength region $\bar{I}_\lambda(T) = a + b(\lambda - \lambda_0)$, λ_0 being some wavelength in the interval $(\lambda_2 - \lambda_1)$, we get

$$\lambda_i = \lambda_0 = \frac{\int_{\lambda_1}^{\lambda_2} \lambda\, s(\lambda)\, d\lambda}{\int_{\lambda_1}^{\lambda_2} s(\lambda)\, d\lambda} = \lambda \qquad (42')$$

and $a = I_{\lambda_i}(T)$.

The isophotic wavelength is sometimes defined in terms of stellar magnitudes. Let m be the integrated stellar magnitude in a particular photometric system, m_λ the monochromatic magnitude, and m_i the monochromatic magnitude at the isophotic wavelength. Then

$$10^{-0.4\,(m-\text{const})} = \int_0^\infty 10^{-0.4 m_\lambda} s(\lambda)\, d\lambda = 10^{-0.4 m_{\lambda_i}} \int_0^\infty s(\lambda)\, d\lambda. \qquad (42'')$$

The problem of determining various wavelengths of photometric and colorimetric systems reduces to the following: first we find the particular wavelength of the entire system which, if used to calculate m_λ, M_λ, C_λ, λ, etc., from the above expressions, gives the same results as the averaged values (calculated making use of the function $s(\lambda)$). In these calculations one should naturally consider the energy distribution in the continuum and the extinction of light in the Earth's atmosphere.

The calculation of various working wavelengths for a given system of magnitudes and colors is of the greatest importance, and we will now consider it in some detail. In our discussion we will follow the treatment of Seares and Joyner (1943). Planck's equation for the outgoing radiation flux (in the observer's hemisphere) in a wavelength interval $\Delta\lambda = 1$ cm is conveniently written in the form

$$H_\lambda = \pi J_\lambda(T) = \pi \frac{C_1}{\lambda^5} \frac{1}{e^{\frac{c_2}{\lambda T}} - 1} \text{ (erg/sec} \cdot \text{cm}^2 \text{ or W/cm}^2\text{)}. \qquad (43)$$

Substituting $x_1 = \frac{\lambda T}{c_2}$, we obtain for a wavelength interval $\Delta\lambda = 1$ cm (i.e., $\Delta x_1 = \frac{T}{c_2} \cdot 1$ cm)

$$H_\lambda = \pi J_\lambda(T) = \pi \frac{C_1}{c_2^5} T^5 \frac{x_1^{-5}}{e^{\frac{1}{x_1}} - 1} = \pi c_3 T^5 y_1 \text{ (erg/sec} \cdot \text{cm}^2 \text{ or W/cm}^2\text{)}. \qquad (44)$$

The constants, as before, are

$$C_1 = 2hc^2 = 1.191 \cdot 10^{-12} \text{ W} \cdot \text{cm}^2/\text{sterad},$$

$$c_2 = \frac{hc}{k} = 1.4385 \text{ cm} \cdot \text{g},$$

$$c_3 = \frac{C_1}{c_2^5} = \frac{2hc^2}{\left(\frac{hc}{k}\right)^5} = 2\frac{k^5}{c^3h^4} = 1.933 \cdot 10^{-13}\ \mathrm{W/cm^3 \cdot g^5 \cdot sterad}.$$

We see that the expression reduces to a product of two functions, T^5 and $y_1 = \frac{x_1^{-5}}{\frac{1}{e^{x_1}-1}}$. The latter function, which is independent of experimental constants, is graphically shown in Figure 2. Its maximum (y_m) is numerically equal to 20.202 for $x_m = 0.20141$ ($\frac{1}{x_m} = 4.9651$). Making another substitution $x = \frac{x_1}{x_m} = \frac{\lambda}{\lambda_m}$ and $y = \frac{y_1}{y_m}$, we obtain for $\lambda_m T = c_2 x_m = 0.2897$ in a wavelength interval $\Delta\lambda = 1$ cm (i.e., for $\Delta x = \frac{\Delta x_1}{x_m} = \frac{1\mathrm{cm} \cdot T}{c_2 x_m} = \frac{1\mathrm{cm} \cdot T}{0.2897\mathrm{cm} \cdot \mathrm{g}}$)

$$H_\lambda = \pi J_\lambda(T) = \pi c_3 T^5 y y_m = \pi c_4 T^5 y = 1.287 \cdot 10^{-11} T^5 y\ \ (\mathrm{erg/sec \cdot cm^2\ or\ W/cm^2}).$$

Here $c_3 y_m = c_4 = 4.099 \cdot 10^{-12}$ and

$$y = \frac{y_1}{y_m} = \frac{1}{y_m} \frac{x_1^{-5}}{\frac{1}{e^{x_1}-1}} = \frac{x_m^{-5}}{y_m} \frac{x^{-5}}{\frac{1}{e^{x_m x}-1}} = \frac{x_m^{-5}}{y_m} \frac{x^{-5}}{\frac{c_5}{e^x-1}}. \qquad (45)$$

Also $\frac{x_m^{-5}}{y_m} = c_6 = 142.32$ and $c_5 = \frac{1}{x_m} = 4.9651$.

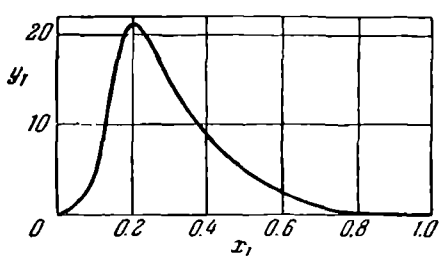


FIGURE 2

Figure 3 plots the function $y = f(x)$ according to Waldmeier (1948). It is normalized and its maximum therefore corresponds to the coordinates $y = 1$, $x = 1$. For a blackbody source, the visual region, i.e., $\lambda = (3.8-7.6) \cdot 10^{-5}$ cm, corresponds at $T = 20{,}000°$K to $x = 2.63-5.26$. These abscissas are marked on top by letters B and C, and are joined by straight lines with the origin. The values of x for the same visual spectrum but blackbody radiation temperatures $T < 20{,}000°$K are determined by the intersection of the lines OB, OC with the ordinates $T/20{,}000$. The cross-hatched rectangles in the figure correspond to $T = 2000°$, $6000°$, $15{,}000°$K for $\lambda = (3.8-7.6) \cdot 10^{-5}$ cm, respectively. The third line OA corresponds to the wavelength $\lambda = 3 \cdot 10^{-5}$ cm (3000 Å) and limits in the left part of the figure the region which is inaccessible to earthbound spectroscopic observations due to atmospheric extinction.

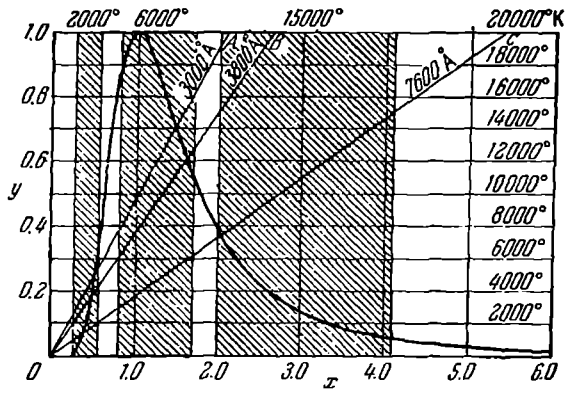


FIGURE 3

Table 1 lists the values of the function y for various temperatures and wavelengths. The table is based on the old value of the constant $\lambda_m T = 0.2886$, which corresponds to $c_2' = 1.433$. The new value, 0.2897, used in the preceding corresponds to $c_2 = 1.4385$ cm · g.

TABLE 1

λ (cm)·10⁴ \ T°K	3000	3500	4000	5000	6000	7000	9000	11000	13700	17000	22000
0.30	0.0059	0.0263	0.0743	0.2667	0.5263	0.7608	0.9885	0.9588	0.7688	0.5308	0.2919
.31	.0088	.0347	.0928	.3083	.5776	.8046	.9969	.9380	.7333	.4963	.2682
.32	.0115	.0448	.1135	.3510	.6269	.8440	.9999	.9143	.6981	.4639	.2466
.33	.0155	.0566	.1367	.3946	.6740	.8787	.9982	.8885	.6637	.4340	.2271
.34	.0204	.0702	.1619	.4390	.7185	.9086	.9922	.8615	.6303	.4056	.2094
.35	.0264	.0856	.1895	.4831	.7598	.9340	.9825	.8335	.5980	.3788	.1931
.36	.0335	.1029	.2188	.5268	.7977	.9549	.9697	.8049	.5672	.3544	.1783
.37	.0418	.1220	.2496	.5696	.8323	.9713	.9548	.7760	.5376	.3316	.1647
.38	.0514	.1429	.2818	.6112	.8636	.9836	.9376	.7473	.5093	.3104	.1526
.39	.0623	.1654	.3152	.6513	.8914	.9923	.9182	.7188	.4826	.2906	.1416
.40	.0746	.1895	.3495	.6896	.9157	.9977	.8976	.6905	.4569	.2722	.1315
.41	.0883	.2151	.3845	.7260	.9367	.9999	.8760	.6629	.4327	.2551	.1222
.42	.1032	.2419	.4197	.7602	.9544	.9992	.5838	.6359	.4101	.2392	.1135
.43	.1195	.2697	.4551	.7921	.9688	.9957	.8308	.6098	.3883	.2246	.1056
.44	.1372	.2985	.4903	.8217	.9802	.9898	.8075	.5844	.3678	.2111	.0985
.45	.1560	.3282	.5253	.8489	.9889	.9817	.7839	.5597	.3482	.1982	.0919
.46	.1761	.3584	.5596	.8737	.9950	.9719	.7604	.5361	.3301	.1863	.0858
.47	.1973	.3890	.5931	.8961	.9986	.9608	.7370	.5134	.3132	.1753	.0801
.48	.2195	.4198	.6257	.9161	.9999	.9483	.7139	.4915	.2969	.1650	.0750
.49	.2425	.4508	.6576	.9337	.9992	.9343	.6909	.4705	.2817	.1556	.0702
.50	.2663	.4818	.6881	.9491	.9965	.9191	.6680	.4502	.2672	.1469	.0658
.51	.2908	.5125	.7174	.9622	.9919	.9032	.6457	.4309	.2537	.1385	.0618
.52	.3160	.5427	.7452	.9730	.9857	.8867	.6241	.4123	.2410	.1307	.0580
.53	.3417	.5724	.7719	.9819	.9780	.8696	.6031	.3948	.2288	.1235	.0545
.54	.3678	.6015	.7968	.9890	.9692	.8521	.5827	.3779	.2175	.1168	.0513
.55	.3941	.6301	.8204	.9943	.9595	.8342	.5628	.3618	.2068	.1104	.0483
.56	.4206	.6579	.8424	.9979	.9488	.8161	.5432	.3465	.1968	.1044	.0455
.57	.4472	.6847	.8629	.9997	.9369	.7978	.5242	.3317	.1873	.0989	.0429
.58	.4737	.7105	.8818	.9999	.9241	.7794	.5059	.3178	.1781	.0937	.0405
.59	.5001	.5353	.8993	.9988	.9107	.7611	.4881	.3046	.1695	.0889	.0383
0.60	0.5263	0.7590	0.9151	0.9963	0.8969	0.7429	0.4710	0.2916	0.1615	0.0843	0.0363

Note that since

$$\pi J_{max}(T) \equiv \pi J_m(T) = H_{max}(T) = \pi c_4 T^5,$$

the transformed and normalized Planck's function has the form

$$\frac{H_\lambda}{H_{\max}}=\frac{J_\lambda}{J_{\max}}=y=f(x)=\frac{x_m^{-5}}{y_m}\frac{x^{-5}}{e^{\frac{c_5}{x}}-1}=c_0\frac{x^{-5}}{e^{\frac{c_5}{x}}-1}. \tag{46}$$

The normalized Planck's function thus coincides with the nondimensional function y presented in Figure 3 and Table 1. It does not depend explicitly on temperature, a very convenient feature for tabulation.

Let us now return to our fundamental problem.

If the star is a blackbody source of temperature T, the measured energy can be written in the form (we take $s(\lambda)=s$)

$$i=T'\int ys\,d\lambda \quad \text{or} \quad m^*=-2.5\lg i=-2.5\lg T'\int ys\,d\lambda. \tag{47}$$

(Here dx is replaced with $d\lambda$, since $x\sim\lambda$). For two wavelengths, photographic and visual, we have the color index

$$C_{p-v}=-2.5\lg\int s_p y\,d\lambda+2.5\lg\int s_v y\,d\lambda. \tag{48}$$

The integrated magnitudes m_p and m_v can be characterized in terms of the effective weighted average wavelengths (equation (39))

$$\bar\lambda_p=\lambda_{e,p}=\lambda_p=\frac{\int\lambda y s_p\,d\lambda}{\int y s_p\,d\lambda},\quad \bar\lambda_v=\lambda_{e,v}=\lambda_v=\frac{\int\lambda y s_v\,d\lambda}{\int y s_v\,d\lambda}. \tag{49}$$

In practice integration is replaced by summation ($\sum\lambda y s_p$ instead of $\int\lambda y s_p\,d\lambda$, etc.). If the total and the observed energy i_p and i_v are expressed in terms of the blackbody monochromatic energy at λ_p and λ_v,

$$i_p=k_p\pi J_p(T),\quad i_v=k_v\cdot\pi J_v(T), \tag{50}$$

we establish a relation between the average wavelengths, on the one hand, and C_{p-v} and T on the other. Here k are coefficients introduced to ensure proper matching in equations.

Since $\pi J_\lambda(T)=T'y$ (equation (49)) and $i=T'\int ys\,d\lambda$ (equation (47)), we may write

$$k_p=\frac{\int y s_p\,d\lambda}{y_p},\quad k_v=\frac{\int y s_v d\lambda}{y_v}. \tag{51}$$

The temperature enters implicitly (through y) the color C_{p-v}, the weighted average wavelength λ_e, and the coefficient k. However, the variation of k with temperature is slow and in general insignificant. Nevertheless, this variation is in practice highly undesirable and it is avoided by changing over to the Brill's isophotic wavelength λ_i (1929) and appropriately modifying the weighted average wavelength of the system $\bar\lambda_E$. We may thus write

$$k y_i=\int ys(\lambda)\,d\lambda. \tag{52}$$

• Note that in practice the apparent stellar magnitudes are used on the scale of c_m with the zero point c_0. Thus $m=-c_m\lg i+c_0$. For the usual Pogson scale $c_m=2.5$ (although a better choice is $c_m=\ln 10=2.303$). In our case, the zero point is chosen so that $c_0=0$.

Here k is independent of T: it depends on $s(y)$ only; y_i is selected from Table 1 for the isophotic wavelength λ_i.

The maximum value of the normalized function y_i is 1. We see from Table 1 that in the working spectrum it is reached at temperatures 5000—10,000°K. Then

$$k = \left[\int ys(\lambda)\, d\lambda\right]_{T_0}^{\max} = \left[\int ys(\lambda)\, d\lambda\right]_{T_0}, \tag{53}$$

and therefore in general

$$y_i = \frac{\left[\int ys(\lambda)\, d\lambda\right]_T}{\left[\int ys(\lambda)\, d\lambda\right]_{T_0}}. \tag{54}$$

Equation (54) enables us to compute y_i for various T and hence to determine the isophotic wavelengths of the photometric or the colorimetric system.

Eqs. (44) and (47) give

$$\frac{J_i(T)}{J_i(T_0)} = \frac{i(T)}{i(T_0)},$$

$$-2.5 \lg \frac{J_i(T)}{J_i(T_0)} = -2.5 \lg \frac{i(T)}{i(T_0)} = m_i'(T) - m_i'(T_0) = m(T) - m(T_0). \tag{55}$$

Here m_i' are the stellar magnitudes computed for the isophotic wavelength λ_i, and m are the apparent stellar magnitudes.

By (55) we have

$$m_i'(T) = m(T) + [m_i'(T_0) - m(T_0)] = m(T) + \text{const}. \tag{56}$$

This result can be formulated as follows: at any blackbody temperature the monochromatic stellar magnitude for isophotic wavelength λ_i is equal to the integrated stellar magnitude ($-2.5 \lg T' \int ys(\lambda)\, d\lambda$) of the blackbody source at the same temperature plus a temperature-independent constant, since T_0 = const and $y_i(T_0) = 1$ for the given system.

We can now write the expression for the color index, i.e., the difference between the stellar magnitudes at two isophotic wavelengths of the photometric system. For example, for the p, v system with isophotic wavelengths λ_{ip} and λ_{iv}, we have

$$C_{p-v} = -2.5 \lg y_{ip} + 2.5 \lg y_{iv} + \text{const}. \tag{57}$$

This equation follows from (48) remembering that k is constant. Using Table 1 and Eq. (56), Seares and Joyner (1943) calculated the λ_i in the international *IPg*, *IPv* system of photographic and photovisual magnitudes. The results are listed in Table 2 for color temperatures from 30,000 to 3000°K.

TABLE 2

T_c °K	30000	22000	17000	13700	11000	9000	7000	6000	5000	4000	3500	3000
λ_{ip}	—	4200	4210	4219	4231	4248	4270	4296	4913	4355	4380	4411
λ_{iv}		5424	5426	5427	5428	5431	5435	5439	5443	5445	5448	5459
$\lambda_{e,p}$	4163	4182	4198	4216	4242	4272	4318	4351	4398	4465	4508	4559
$\lambda_{e,v}$	5415	5418	5420	5423	5426	5430	5437	5442	5449	5460	5468	5478

The international system of stellar magnitudes and its zero point are defined in terms of the North Polar Sequence stars with photovisual magnitudes $m_{pv} = +2^m.08 - 17^m.47$ and photographic magnitudes up to $m_{pg} = +20^m.18$. The table lists for comparison the weighted average wavelengths (sometimes called effective by mistake).

The variation of isophotic wavelengths with temperature is approximately half the variation of λ_e (the weighted average wavelengths), which were previously measured with crude objective diffraction gratings. Therefore $\overline{\lambda}_e = \lambda_e$ provide a more appropriate characteristic of the stellar radiation.

If the color indices in a given colorimetric system are compared with the international system by means of the so-called color equation, Eq. (57) and Table 1 can be used to determine the λ_i for this system.

Unfortunately stars are not blackbody sources; in the best case the energy distribution in the continuum corresponds to the spectrum of a gray body. The outgoing radiation flux of a real star is therefore given by

$$H_\lambda^0 = g\pi J_\lambda (T_e). \tag{58}$$

Here the factor g characterizes the extent to which the particular star departs from a blackbody. In the case of an ideal gray body this factor is independent of both T_e and λ.

Just outside the Earth's atmosphere* we can in principle measure the "extra-atmospheric" monochromatic radiation (illuminance)

$$e_\lambda = H_\lambda^0 \frac{R^2}{r^2} = g\pi J_\lambda (T_e) \frac{R^2}{r^2}, \tag{59}$$

and the visual stellar magnitude m_v (or the photovisual m_{pv}) is given by

$$m_v = m_v' - 2.5 \lg \frac{e_\lambda}{e_\lambda'} \tag{60}$$

or, using the factor $s(\lambda)$ (instrumental and atmospheric transmittance)

$$m_v = m_v' - 2.5 \left\{ \lg \int e_\lambda s(\lambda)\, d\lambda - \lg \int e_\lambda' s(\lambda)\, d\lambda \right\}. \tag{61}$$

By Eq. (40),

$$\int J_\lambda (T_e)\, s(\lambda)\, d\lambda = k J_{\lambda_i} (T_e)$$

where

$$k = \frac{\int J_\lambda (T_0)\, s(\lambda)\, d\lambda}{J_{\overline{\lambda}_E} (T_0)}.$$

We may thus write

$$\int e_\lambda s(\lambda)\, d\lambda = g\pi J_{\lambda_i} (T_e) \frac{R^2}{r^2}. \tag{62}$$

Further, if λ_i is close to $\overline{\lambda}_E$, then following Woolley and Stibbs (1953) we may write, as usual,

$$J_{\lambda_i} (T_0) = J(\lambda_0, T_e) + (\lambda_i - \overline{\lambda}_E) \left\{ \frac{\partial J(\lambda, T_0)}{\partial \lambda} \right\}_{\overline{\lambda}_E}. \tag{63}$$

* Otherwise, a correction for the atmospheric extinction of light has to be introduced.

Making use of eqs. (60) and (61), we obtain

$$m_{\bar{\lambda}_E} - m'_{\bar{\lambda}_E} = m_v - m'_v + 2.5(\lambda_i - \bar{\lambda}_E)\frac{\partial}{\partial\lambda}\lg J(\lambda, T_c) - 2.5(\lambda'_i - \bar{\lambda}_E)\frac{\partial}{\partial\lambda}\lg J(\lambda, T'_c). \quad (64)$$

Thus, if λ_i is known, this equation can be used for direct conversion of the difference of observed stellar magnitudes in the international photovisual system to monochromatic stellar magnitudes. All systems of stellar magnitudes in photometry and colorimetry are differential systems, with a certain zero point.

§4. Modern color systems. Interstellar absorption. We will briefly touch on this problem as it was not discussed in Vol. I of the Course and is significant for further treatment.

The six-color (U, V, B, G, R, I) photoelectric colorimetric system of Stebbins and Whitford enjoyed considerable popularity in the 1940's. Three- and two-color photoelectric colorimetric systems were used on an even wider scale. These included the three-color (U, B, V) system of Morgan and Johnson,† the two-color m, C (or Pg_e, Pv_e, C_p in the first approximation) system of Eggen, the two-color (R, I) system of Kron and Smith, the two-color system of Stebbins, Whitford, Johnson, and others. The latter, following Kron's suggestion (1951), was called the (P, V) system, since previously various photoelectric systems had received the same designation (Pg_e, C_p), which led to inevitable confusion. This system is based on observation of 9 stars of the North Polar Sequence (Nos. 6, 2r, 10, 4r, 13, 8r, 16, 19, 12r) and, as we see from Table 3, it is close to the international system (IPg, IPv or Pg_{int}, Pv_{int}, C_{int}). This system was recommended at the Pasadena Conference (1950) for photoelectric work at the Lick, Washburn, and Mount Wilson and Palomer observatories. The other systems are therefore generally converted to this particular system; an additional index is introduced in this case. For example, the conversion of Eggen's m, C system gives the magnitudes P_E, V_E and the color index $(P - V)_E$. Comparison of the color indices $(P - V)_E$ and $(B - V)$ and subsequent conversion using the standard formula gives the color indices $(P - V)_{EJ}$, etc. Generally, but not always (only for apparently bright near stars, whose brightness and color are not distorted by galactic absorption), the "normal color" systems are related by linear expressions with some dependence of the parameters on stellar luminosities (giants, dwarfs, etc.). The Lick three-color system is $P \approx IPg, V = IPv, I_{IR} \approx I$ (Kron, $\bar{\lambda}_e = 8250$ Å).

TABLE 3

NPS	$P=Pg_p$	$Pg_{int}=IPg$ **	$P-Pg_{int}$	$(P-V)=C_p$	C_{int}*	$(P-V)-C_{int}$	Sp (HD)	$V=Pv_p=Pg_p-C_p$	$IPv=IPg-C_{int}$
6	$7^m.15$	$7^m.12$	$+0^m.03$	$+0^m.06$	$+0^m.06$	$0^m.00$	A2	$7^m.09$	$7^m.06$
2r	7 .90	7 .91	−0 .01	+1 .56	+1 .57	−0 .01	gM2	6 .34	6 .34
10	9 .17	9 .15	+0 .02	+0 .12	+0 .12	0 .00	A6	9 .05	9 .03
4r	9 .24	9 .23	+0 .01	+1 .02	+1 .02	0 .00	gG9	8 .22	8 .21
13	10 .51	10 .55	−0 .04	+0 .24	+0 .22	+0 .02	A5	10 .27	10 .33
8r	11 .43	11 .43	0 .00	+1 .02	+0 .96	+0 .06	gG7	10 .41	10 .47
16	11 .56	11 .57	−0 .01	+0 .32	+0 .34	−0 .02	(F7):	11 .24	11 .23
19	12 .66	12 .68	−0 .02	+0 .41	+0 .44	−0 .03	(F8):	12 .25	12 .24
12r	13 .80	13 .78	+0 .02	+1 .27	+1 .30	−0 .03	—	12 .53	12 .48

* Improved data of Seares and Joyner (ApJ, 101:15. 1945).

** Values of IPg standards from Trans. IAU, 1:69. 1922 (the excess of the new values C_{int} is equally divided between IPg and IPv).

† The stellar magnitude V is close to IP_g, but B is different from IP_v. It is obtained with a blue filter which cuts off all radiation shortward of 3800 Å.

Approximate data on the different systems are listed in Table 4.

TABLE 4

Six-color system, $T=?$, $I_\lambda=$const	U $\bar{\lambda}=3530$ Å	V $\bar{\lambda}=4220$ Å	B $\bar{\lambda}=4880$ Å	G $\bar{\lambda}=5700$ Å	R $\bar{\lambda}=7190$ Å	I (color $V-I$, etc.) $\bar{\lambda}=10\,300$ Å
Three-color system, $T=10\,000$ °K	U $\lambda_i=3560$ Å		B $\lambda_i=4400$ Å		V (color $U-B$, $B-V$) $\lambda_i=5530$ Å	
Two-color system, $T=?$, $I_\lambda=$const	m, Pg_e $\bar{\lambda}_E=4150$ Å			m, $Pv_e=Pg-C$ (color C, C_p) $\bar{\lambda}_E=5100$ Å		
Two-color system, $T=?$, $I_\lambda=$const	P $\bar{\lambda}=4215$ Å			V (color $P-V$) $\bar{\lambda}=5240$ Å		
International two-color system, $T=11\,000$ °K	IPg $\bar{\lambda}_E=4240$ Å			IPv $\bar{\lambda}_E=5430$ Å		
Two-color system, $T=?$, $I_\lambda=$const	I $\bar{\lambda}_E=8250$ Å			R $\bar{\lambda}_E=6500$ Å		

It should be kept in mind that the effective wavelengths are defined differently in different cases.* For example, in the six-color system (U, V, B, G, R, I) the weighted average wavelengths refer to a source of uniform brightness over the entire spectrum. They are thus different from the ordinary — average, effective, and isophotic — wavelengths, which incorporate the energy distribution in the continuum, differential atmospheric extinction, losses in telescopic optics, detector response, etc. The values of $\bar{\lambda}$ listed in the table for the six-color photometry apply to filter + detector combinations and therefore depend on the color index of the star. In photoelectric colorimetry this effect is generally negligible compared, say, to the same effect in photographic colorimetry. This is due, in particular, to the fact that in photoelectric observations the actual working region of the spectrum is mostly narrower than in photographic observations and the shift of the weighted average or effective wavelengths with temperature is correspondingly smaller. Thus, according to Seares and Joyner (1943), a change in temperature from 3000 to 30,000°K shifts the photographic effective wavelengths of the International System in the short-wave direction by 396 Å from 4554 Å.** At the same time, the photoelectric effective wavelength of Stebbins's system C_1 ($\bar{\lambda}_1=4190$ Å, $\bar{\lambda}_2=4760$ Å, $C_1=0.144+0.483\,(B-V)$) shifts by −228 Å and −253 Å from 4693 Å and 4629 Å, respectively, i.e., nearly half as much.

* Recently, Mel'nikov obtained the following isophotic wavelengths for the $(B-V)$ system:

T, °K	3000	6000	10000	15000	25000	50000
λ_i (B)	4490	4410	4380	4370	4360	4350
λ_i (V)	5540	5480	5470	5460	5460	5460

** In the B, V system the effective λ_e^B shifts in the short-wave direction by 270 Å from 4610 Å and the isophotic λ_i^B shifts by 135 Å from 4490 Å (author's data, see footnote above).

For the systems (m, C), (Pg_e, Pv_e), (P, V) and (R, I) Table 4 also gives the weighted-average wavelengths. Conversely, for the (U, B, V) system the isophotic wavelengths are given, as calculated by K. Schmidt (1956) for a blackbody with $T = 10{,}000°K$. For the classical international system IPg, IPv the weighted averaged wavelengths are given for a blackbody with $T = 11{,}000°K$.

The zero point of the popular $(U\ B, V)$ system is defined for $U-B = B-V = 0$ for six A0V stars (MK): α Lyr, γ UMa, 109 Vir, α Cr B, γ Oph and HR 3314. Moreover, it is additionally obtained

1) that for K0—A0, $\Delta(B-V) = \Delta(U-B) = 1$, i.e., these color indices change by unity on passing from A0 to K0,

2) that $V = IPv + \underset{\pm\ 6}{0.000} + \underset{\pm\ 5}{0.002}\ (B-V)$.

It should be kept in mind, however, that the wavelengths $\overline{\lambda}_E$ in the international system are partially distorted by atmospheric absorption at the height of the North Pole (at the Mount Wilson Observatory). No correction is generally introduced for this absorption. The new systems described above, with the exception of the International System, are all photoelectric. This also applies to the high-precison system of V. B. Nikonov which uses a large basis $(\lambda_2 - \lambda_1)$. The U, B, V system was adopted at the Dublin IAU Conference (1955) as the new international system. It had been commonly accepted in photographic observations in the USA and Europe. For example, K. Schmidt (1956) used Schott filters and Agfa Wolfen photographic plates to obtain the isophotic wavelengths 356, 433, and 452 mμ ($T = 10{,}000°K$) for the U, B, V system, which virtually coincide with the original photoelectric isophotic wavelengths (Table 4).

Figure 4 gives a graphic comparison of the color indices in the systems $C = V - B$ and $C = R - I$ according to Kron's recent observations (1958). The thick line and the circles define the true colors of stars of various spectral types, the thin parallel arrows correspond to the colors obtained with allowance for interstellar absorption (see below). Figure 5 gives an analogous comparison for the systems $C=B-V$ and $C = P - V$. In this case a linear relation is observed. Note that interstellar absorption by dust particles makes the color of white stars, say, yellow or even red. The observations thus give an energy continuum which corresponds to a lower color temperature. Let us calculate this hypothetical (apparent) lowering of the star temperature.

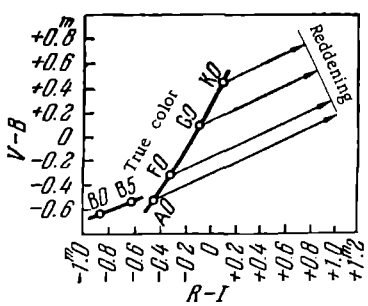


FIGURE 4

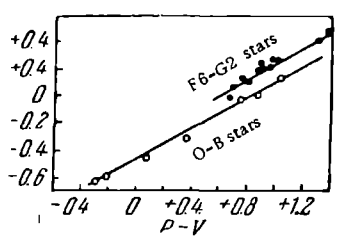


FIGURE 5

Without interstellar absorption, we have for a blackbody source of temperature T in Wien's approximation* for photovisual wavelengths, say,

$$J_\lambda(T) = \frac{C_1}{\lambda_{pv}^5} e^{-\frac{c_2}{\lambda_{pv} T}}. \tag{65}$$

Conversely, absorption by a homogeneous layer of optical thickness $\tau(\lambda_{pv}) \approx \frac{\tau}{\lambda_{pv}}$, which is proportional to the distance, gives

$$J'_{\lambda_{pv}} = J_{\lambda_{pv}}(T)\, e^{-\tau(\lambda_{pv})} = \frac{C_1}{\lambda_{pv}^5} e^{-\frac{c_2}{\lambda_{pv} T}} e^{-\frac{\tau}{\lambda_{pv}}} = \frac{C_1}{\lambda_{pv}^5} e^{-\frac{1}{\lambda_{pv}}\left(\frac{c_2}{T}+\tau\right)} = \frac{C_1}{\lambda_{pv}^5} e^{-\frac{c_2}{\lambda_{pv} T_\tau}}. \tag{66}$$

Hence the apparent color temperature from continuum measurements

$$T_\tau = \frac{1}{\frac{1}{T} + \frac{\tau}{c_2}} = T\left(1 + \frac{\tau T}{c_2}\right)^{-1} \tag{67}$$

and the corresponding "reddening" is

$$\Delta T = T - T_\tau = T\left[1 - \left(1 + \frac{\tau T}{c_2}\right)^{-1}\right]. \tag{68}$$

By observing two stars, one far away with its continuum greatly attenuated by interstellar absorption and another nearby, for which the absorption is smaller, we can easily calculate the difference of their apparent magnitudes. If R_1, r_1, T_1 and R_2, r_2, T_2 are the radii, distances, and temperatures of the two stars and $\tau(\lambda) = \tau_0 + \frac{\tau}{\lambda^\alpha}$ is the optical thickness of the homogeneous interstellar medium, we obtain in the Wien's approximation for the illuminances $e_1(\lambda)$ and $e_2(\lambda)$ at the detector (ignoring atmospheric absorption and other corrections)

$$\left.\begin{aligned} e_1(\lambda) &\sim \frac{R_1^2}{r_1^2} \lambda^{-5} e^{-\frac{c_2}{\lambda T_1}} e^{-\tau_1(\lambda)}, \\ e_2(\lambda) &\sim \frac{R_2^2}{r_2^2} \lambda^{-5} e^{-\frac{c_2}{\lambda T_2}} e^{-\tau_2(\lambda)}. \end{aligned}\right\} \tag{69}$$

Since $m_1(\lambda) - m_2(\lambda) = -2.5 \lg \frac{e_1(\lambda)}{e_2(\lambda)}$, putting

$$\tau_1(\lambda) - \tau_2(\lambda) \equiv \Delta\tau(\lambda) = \Delta\tau_0 + \frac{\Delta\tau}{\lambda^\alpha},$$

$$A = -5 \lg \frac{R_1}{R_2} + 5 \lg \frac{r_1}{r_2} + 1.086 \Delta\tau_0 = M_1 - M_2 + 10 \lg \frac{T_1}{T_2} + 1.086 \Delta\tau_0,$$

$$B = 1.086 c_2 \left(\frac{1}{T_1} - \frac{1}{T_2}\right), \quad C = 1.086 \Delta\tau,$$

we finally obtain (for $\alpha = \text{const}$ in a narrow spectral interval)

$$\Delta m(\lambda) = m_1(\lambda) - m_2(\lambda) = A + \frac{B}{\lambda} + \frac{C}{\lambda^\alpha}. \tag{70}$$

* To within 1%, Wien's approximation to Planck's equation can be used when $\left(\frac{c_2}{\lambda T}\right) > 100$ or $\lambda T < 0.31$; since $\lambda_m T = 0.29$, this means that $\lambda < 1.07 \lambda_m$. Wien's approximation is thus applicable to within 1% shortward of the continuum maximum of the given star (on the wavelength scale).

If the two comparison stars are of the same spectral type, $B = 0$. The interstellar extinction $\Delta m(\lambda)$ is generally denoted by $A(\lambda) = a(\lambda) r$. Here r is the distance of the star in kiloparsecs, and $a(\lambda)$ is the interstellar extinction coefficient in stellar magnitudes per 1 kiloparsec. The coefficient $\varkappa(\lambda)$ introduced in what follows is a volume absorption coefficient and as in physics it is calculated per 1 cm.

By definition, for a homogeneous absorbing layer of optical thickness $\tau(\lambda) = \varkappa(\lambda) r$, we have

$$A(\lambda) = -2.5 \lg e^{-\tau(\lambda)} = 1.086\,\tau(\lambda) = 1.086\varkappa(\lambda) r\,(\text{cm}) = 3.35 \cdot 10^{25}\varkappa(\lambda) r\,(\text{kpc}), \quad (71)$$

and the stellar coefficient is thus related to the physical coefficient by

$$a(\lambda) = 3.35 \cdot 10^{25}\varkappa(\lambda). \quad (72)$$

Since in the Galaxy we on the average have for various latitudes and longitudes $a(\lambda_{pv}) \approx 1\ \text{kpc}^{-1}$, we find $\varkappa(\lambda_{pv}) \approx 3 \cdot 10^{-22}$. The observed stellar magnitudes should be corrected for interstellar absorption. This, however, is one of the most difficult problems of modern astrophysics. The correction is generally introduced using the expressions

$$\left.\begin{aligned} A(\lambda_{pg}) &\equiv A(\lambda_p) = \gamma_p [A(\lambda_p) - A(\lambda_v)] \equiv \gamma_p E, \\ A(\lambda_{pv}) &\equiv A(\lambda_v) = \gamma_v [A(\lambda_p) - A(\lambda_v)] \equiv \gamma_v E. \end{aligned}\right\} \quad (73)$$

Here γ_p and $\gamma_v \approx$ const; according to Mel'nikov (1944), in the International System $\gamma_{IPg} = 4.6$, $\gamma_{IPv} = 3.6$.

E is readily obtained from observations (this is the so-called color excess; $\varepsilon = \frac{E}{r}$ is the coefficient of selective absorption of the homogeneous medium) by comparing the apparent color index with the true color index for the given temperature. The coefficient $\gamma_p = 1 + \gamma_v$ can be calculated if the interstellar extinction characteristic is known. We may take $\gamma_p = 4.6$, $\gamma_v = 3.6$ (Mel'nikov, 1944). According to the author (1958), the best approximate formula is

$$A(\lambda) = A_\lambda + A_0 = \frac{A(0)}{1 + A\lambda^3} + A_0. \quad (74)$$

Here A_0 is the extinction component which is independent of wavelength, i.e., neutral extinction ($A(\lambda) \to A_0$ for $\lambda \to \infty$) and A_λ is the selective extinction ($A_\lambda \to A(0)$ for $\lambda \to 0$). When working in narrow spectral intervals, one may use the old relation (as in Eq. (70)) $A(\lambda) = A_\lambda + A_0 = \frac{\mathfrak{A}}{\lambda^\alpha} + A_0$.

In particular, for the photographic spectrum $\alpha \approx 1$ and the selective extinction $\sim \lambda^{-1}$. The values of α for other spectral regions were tabulated by the author (1944).

§5. Radiation fluxes (illuminances) and stellar brightness. Observations of monochromatic radiation fluxes corrected for interstellar extinction can be used in quantitative analysis of radiation characteristics. A significant step in the whole procedure is the conversion of the observed radiation fluxes (illuminances) to true fluxes per unit surface. This can be done by comparing the system of the stellar magnitude with an earthbound laboratory source, or, even simpler, with the Sun.

The photovisual magnitude of the Sun in the international system ($\lambda_{pv} = 5430$ Å) is $26^m.84$ (Kuiper, 1938, Martynov, 1960, and others). The photovisual flux radiated by every square centimeter of the Sun's surface into a hemisphere is (according to N. Milford, normalized and corrected for the effect of lines and bands according to Wempe, 1947 and Michard, 1950)

$$H_{pv} = \pi \bar{I}_{pv} = \pi \cdot 3.21 \cdot 10^{14} = 10.1 \cdot 10^{14} \text{ erg/sec} \cdot \text{cm}^2 = 10.1 \cdot 10^7 \text{ W/cm}^2.$$

Here, as before $\Delta\lambda = 1$ cm. The luminosity of the Sun in the photovisual spectrum is correspondingly

$$(L_\odot)_{pv} = 4\pi R_\odot^2 H_{pv} = L_{pv,\odot}.$$

The mean angular radius of the Sun at a distance $r_\odot = 4.85 \cdot 10^{-6}$ pc (the parallax is $p_\odot = \frac{1}{4.85 \cdot 10^{-6}} = 2''.063 \cdot 10^5$) is

$$\frac{R_\odot}{r_\odot} = \frac{1919''}{2} = 959''.5 \approx 16' = \frac{6.953 \cdot 10^{10} \text{ cm}}{1.495 \cdot 10^{13} \text{ cm}} = 0.004652 \text{ radian.} \qquad (75)$$

The illuminance produced by the Sun on a unit surface area just outside the Earth's atmosphere in the photovisual spectrum with $\Delta\lambda = 1$ cm is thus

$$e_{pv,\odot} = \frac{L_{pv,\odot}}{4\pi r_\odot^2} = H_{pv}\left(\frac{R_\odot}{r_\odot}\right)^2 = \pi \bar{I}_{pv}\left(\frac{R_\odot}{r_\odot}\right)^2 = \pi \cdot 3.21 \cdot 10^{14} \cdot (0.004652)^2 =$$
$$= \pi \cdot 6.85 \cdot 10^9 = 2.18 \cdot 10^{10} \text{ erg/sec} \cdot \text{cm}^2 = 2.18 \cdot 10^3 \text{ W/cm}^2. \qquad (76)$$

The illuminance of a unit surface area produced by a star of magnitude m — a dwarf of spectral class G2V (according to the latest data, the Sun is a star of this type) — is thus

$$e_{pv,m} = e_{v,m} = e_{v,\odot} 10^{-0.4(m_v - m_\odot)}.$$

For $e_{v,\odot} = 2.18 \cdot 10^{10}$ erg/sec · cm^2, $m_v = -26^m.84$, $\lambda_v = 5430$ Å, $\Delta\lambda = 1$ cm, and $m = 0^m.0$, we have

$$e_{v,0} = 2.18 \cdot 10^{10} \cdot 10^{-10.736} = 0.40 \text{ erg/sec} \cdot \text{cm}^2 = 4.0 \cdot 10^{-6} \text{ W/cm}^2. \qquad (77)$$

In general for a m mag. star ($\Delta\lambda = 1$ cm)

$$\left.\begin{aligned} \lg e_{v,m} \text{ (erg/sec} \cdot \text{cm}^2) &= \lg e_{v,\odot} - 0.4(m_v - m_{v,\odot}) = -0.40 - 0.40 m_v, \\ \lg e_{v,m} \text{ (W/cm}^2) &= -7.40 - 0.40 m_v. \end{aligned}\right\} \qquad (78)$$

For $\Delta\lambda = 1$ Å, i.e., at $\lambda = 5430$ Å between 5429.5 and 5430.5 Å, we have in absolute units

$$\left.\begin{aligned} \lg e_{v,m} \text{ (erg/sec} \cdot \text{cm}^2) &= -8.40 - 0.4 m_v, \\ \lg e_{v,m} \text{ (W/cm}^2) &= -15.40 - 0.4 m_v. \end{aligned}\right\} \qquad (79)$$

Eq. (64) can be used for conversion to stars of other spectral classes. Let the isophotic wavelength λ_i and the color temperature T_c correspond to a star of an arbitrary spectral type, and λ_i', T_c' to a G2V dwarf; then in a 100 Å interval we have

$$\lg e_\lambda \text{(erg/sec} \cdot \text{cm}^2) = -6.40 - 0.4 m_v - (\lambda_i - \bar{\lambda}_E) \frac{\partial}{\partial \lambda} \lg J_\lambda(T_c) + (\lambda_i' - \bar{\lambda}_E) \frac{\partial}{\partial \lambda} J_\lambda(T_c'). \qquad (80)$$

Here $\bar{\lambda}_E = 5430$ Å. The correction $\lg(\lambda_i - \bar{\lambda}_E)\frac{\partial}{\partial\lambda}\lg J_\lambda(T_c)$ is small (except for very cold stars). For $T = 22{,}000$, 14,300, 6750, and 3000°K it is a mere $+0^m.00156$, $+0^m.00068$, $-0^m.00048$ and $+0^m.00728$, respectively, i.e., less than the accuracy of the international stellar magnitudes ($\pm 0^m.01$).

To convert e_v to $H_v = \pi\bar{I}_v = \left(\frac{r}{R}\right)^2 e_v$ for

$$\lg H_{v,\odot} = \lg H_{5430}\,(\mathrm{erg/sec\cdot cm^2}) = \lg(\pi\cdot 3.21\cdot 10^{14}) = 15.0\,(\Delta\lambda = 1\,\mathrm{cm}) = 7.0 \qquad (\Delta\lambda = 1\,\text{Å})$$

we require M or R, r or p''(the parallax). For the bolometric fluxes we have

$$\lg\frac{H_b}{H_{b,\odot}} \equiv \lg\frac{H}{H_\odot} = 2\lg\left(\frac{p_\odot R_\odot}{pR}\right) - \frac{1}{2.5}(m_b - m_{b,\odot}). \tag{81}$$

Using the Stefan—Boltzmann law $H = \sigma T^4$ for visual (more precisely, photovisual) fluxes we obtain ($\lambda_{pv} = 5430$ Å)

$$\lg\frac{H_{pv}}{H_{pv,\odot}} \equiv \lg\frac{H_v}{H_{v,\odot}} = 4\lg\frac{T_e}{T_{e,\odot}} + \frac{1}{2.5}(B.C. - B.C._\odot) = 4\lg\frac{T_e}{T_{e,\odot}} - \frac{1}{2.5}(\Delta m_b - \Delta m_{b,\odot}). \tag{82}$$

Here, as before, $\Delta m_b = m_v - m_b = -B.C.$ is the bolometric correction. The bolometric magnitudes m_b should be corrected for the effect of absorption lines and bands: $B.C'. = -\Delta m_b' - (m_v + \Delta m_v)$.

The data for the bright stars are given in Table 5 following N. Milford (1950). The tabular value of $\lg H_{5430}$, and hence H_{5430}, can easily be converted to e_{5430}. Specifically,

$$e_{pv} = e_v = e_{5430} = H_{5430}\left[\frac{\frac{R}{R_\odot}R_\odot(\mathrm{cm})}{\frac{1}{p''}\cdot 3.084\cdot 10^{18}}\right]^2. \tag{83}$$

This value can be used in the expression relating e_v to m_v.

TABLE 5

Star	Sp	p''	$\frac{R}{R_\odot}$	m_{pv}	Δm_{pv}	$\lg\frac{H_{pv}}{H_{pv,\odot}}$	$\lg H_{5430}$*	$\lg D''$	T_m °K	T_e (Sp) °K
Sun	G2V	$2.063\cdot 10^5$	1.00	−26.84	−0.08	0.00	7.00	+3.28	6 000	5 730
μ^1 Sco A	B3	0.074**	5.26	+3.69	−0.00	+1.21	8.21	−3.44	12 600	18 000**
μ^1 Sco B	B3	0.074**	5.86	+3.98	−0.00	+0.99	7.99	−3.39	10 400	18 000
β Per	B8V	0.032	2.7	+2.3	−0.00	+1.08	8.08	−3.09	12 350	12 800
α CMa	A1V	0.376	1.93	−1.58	−0.01	+0.85	7.85	−2.17	9 830	10 300
β Aur A	A2IV	0.037	2.49	+1.76	−0.01	+0.83	7.83	−3.08	10 500	9 700
β Aur B	A2IV	0.037	2.28	+2.89	−0.01	+0.86	7.86	−3.10	10 300	9 700
α Boo	K2p	0.092	23	+0.04	−0.22	−0.72	6.28	−1.70	4 400	3 810
κ Vir	gK2	0.020	27	+4.3	−0.22	−1.24	5.76	−2.30	3 710	4280-3810
α Tau	K5III	0.059	36	+1.02	−0.30	−1.08	5.93	−1.70	3 900	3 550
YYGemA	K6+	0.073	0.66	+9.57	−0.35	−1.19	5.81	−3.35	3 780	4 000
YYGemB	K6+	0.073	0.60	+9.81	−0.35	−1.21	5.79	3.39	3 760	4 000

* H_{pv} in units erg/sec·cm², $d\lambda = 1$ Å.
** Parallax unreliable.

Recently E. Lamma obtained improved values of the monochromatic flux densities $\lg H_\lambda$ for stars of various types from B0V to M2—M3Iab (1959) in

the region $\lambda = 3000-10{,}000$ Å. To this end, he derived the standard energy distribution of an average A0 star from numerous observations.

The bolometric correction introduced in the preceding should also be calculated for a particular colorimetric system taking into consideration the extinction of the continuous spectrum, the spectral response of the instruments, etc. The expression used to this end is ($L_\lambda \sim \overline{I}_\lambda$)

$$\Delta M_b = \Delta m_b = -B.C. = -2.5 \lg \frac{\int_0^\infty L_\lambda(T)\, s(\lambda)\, d\lambda}{\int_0^\infty L_\lambda(T)\, d\lambda} = -2.5 \lg \frac{\int_0^\infty \overline{I}_\lambda(T)\, s_b(\lambda)\, d\lambda}{\int_0^\infty \overline{I}_\lambda(T)\, d\lambda} = -2.5 \frac{\int y s_b(\lambda)\, d\lambda}{\int y\, d\lambda}. \qquad (84)$$

As we have already noted, the best results for early-type stars are obtained using the continua $\overline{I}_\lambda(T)$ calculated theoretically for the given effective temperature T. These calculations were carried out by various astrophysicists, specifically Mustel (1953) in the USSR and recently Popper (1959) in the West.

TABLE 6

Author	Year	T_e °K	ΔM_b I		ΔM_b IV
E. R. Mustel'	1953	10 500	0^m.45	0^m.72	0^m.68
		12 500	0 .83	1 .11	1 .07
		15 000	1 .24	1 .51	1 .51
		17 500	1 .61	1 .88	1 .87
		20 000	1 .91	2 .19	2 .18
K. Hunger	1954	8160	0.11 II		0.25
		8660	0.14		0.34
		9000	0.18		0.40
		9500	0.24		0.48
G. Traving	1955	32 800	3.20		3.33
J. Pecker	1956	27 300	3.00		2.89
K. Osawa	1956	7 560	0.15 III		0.13
		8 900	0.24		0.39
S. Saito	1956	10 600	0.39		0.69
		15 400	1.22		1.58
		20 500	2.00		2.23

I Two alternatives a) $\Delta M_b = 0$ for $T_e = 6500$°K, $I_\lambda(T)$ Planck's distribution,
b) $I_\lambda(T)$ from the theory of continuous stellar spectra (with the Sun as the zero point).
II ΔM_b somewhat modified by using a different $\Delta M_{b,\odot}$.
III The last five corrections calculated by Popper from $H_\lambda(T)$.
IV Kuiper's corrections calculated from $H_\lambda(T)$.

The results of various authors are listed in Table 6. For comparison the last column gives Kuiper's old data (1938) obtained by the same method for the theoretical H_λ for early-type stars. For the late-type stars (F0—M5) the corrections Δm_b were calculated by Kuiper from radiometric observations. The corrections for types earlier than F0 were computed

from the old theory of stellar continuous spectra. However, Kuiper was the first to allow for the deviation of the stellar radiation from the blackbody spectrum. Note that for hot stars, when $\frac{c^2}{\lambda T} \ll 1$ (the Rayleigh—Jeans approximation), the analog of (35) with $M_v = M_0 + 11.63 - 5 \lg \frac{R}{R_\odot} - 2.5 \lg T$ is written in the form

$$\Delta M_b = -31.55 + 7.5 \lg T. \qquad (85)$$

Considerable improvement of ΔM_b for bright stars is to be expected following direct observation of ultraviolet spectra from rockets and artificial satellites.

Popper (1959) succeeded in somewhat improving the ΔM_v values from existing data. To this end he used 1) radiometric and photoelectric observations of late-type stars in the (B, V) system and 2) an improved model of early-type stellar atmospheres with data on outgoing fluxes H_λ. Unlike the previous authors, he tabulated the corrections according to the color index $(B - V)$, and not according to T or Sp; moreover, for the visual magnitude he used the photoelectric figure V, and not m_{pv} or m_{pg}. The $(B - V)$ color index is close to that in the old international system. The results for stars of later types than F0 are listed in Table 7. The zero point is so chosen that the correction for the main sequence stars G0—G8 coincides with Kuiper's correction (1938). The color index of the Sun in the $(B - V)$ system is $+0^m.63$ and therefore $\Delta M_b = \Delta m_b = 0^m.07$ for the Sun. For early-type stars the corrections were calculated theoretically; they are characterized by the data of Table 5 tabulated according to T_e, since in this case the $(B - V)$ color indices are virtually inapplicable. The new corrections for early-type stars are apparently $0.1 - 0^m.2$ less than Kuiper's old corrections (1938). For O—B -type stars the corrections were derived by Underhill (1957).

TABLE 7

Color index $B-V$	$+0^m.4$	$0^m.5$	$0^m.6$	$0^m.7$	$0^m.8$	$0^m.9$	$1^m.0$	$1^m.1$	$+1^m.2$
$\Delta M_b = -B.C.$	$0^m.00$	$0^m.02$	$0^m.07$	$0^m.11$	$0^m.17$	$0^m.28$	$0^m.40$	$0^m.53$	$0^m.72$

The problem of determination of ΔM_b, Δm_b or $-B.C.$ is closely related to the question of the scale of effective stellar temperatures. Kuiper's scale is currently used, which was introduced in his work on the bolometric correction (1938). The scale was subsequently improved by Morgan using the luminosity classes and the spectral types of stars in the Yerkes systems MKK and MK (1951). Popper's revision (1959) nevertheless led to uncertain results.

Using Eq. (30) for stars whose parallaxes and linear diameters $\frac{D}{D_\odot}$ (eclipsing and spectral binaries) or angular diameters D'' (interferometric measurements) are known, we can find the temperatures $T = T_m$ from the apparent magnitude. The results are listed in Table 5, where the last column also gives the T_e (Sp) scale of 1951. If the luminosity classes are

not known, the absolute magnitudes are computed from the apparent magnitude and the parallax. The values of D'' are based on direct (Sun), interferometric (αCMa, αBoo, etc.) or even spectrophotometric (μ^1Sco, YY Gem, etc.) measurements. Spectrophotometric measurements require knowledge of the exact parallax, light curve, and the spectroscopic orbit of the two components. More or less reliable results are available for three stars only (β Aur, YY Gem, and μ^1Sco).

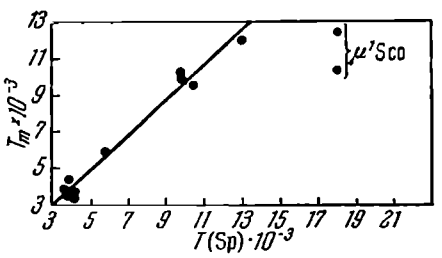


FIGURE 6

T_m is plotted against T(Sp) in Figure 6. We see that the fit is quite satisfactory. The only exception is the μ^1 Sco system, whose parallax is not reliable. The T_e(Sp) scale of 1951 is apparently close to the true scale, but it nevertheless should be further improved. A number of astrophysicists are currently engaged on this problem.

Note that for close binaries the monochromatic fluxes H_{ν} and temperatures can be calculated without the parallaxes, provided the two components have different temperatures and the brightness has been observed in different spectral regions, so that $H_\lambda^{II} - H_\lambda^{I}$ and hence the color indices C can be found. Generally spectroscopic observations of at least one component are available; its color C^I is then determined from the spectrum, and after that the color of the second component C^{II} and its temperature (spectral type) are obtained.

§6. Methods of practical determination of absolute stellar magnitudes. We have so far discussed the uses of absolute magnitudes. Now is clearly the place to describe how they are determined in practice.

1. The simplest method uses the trigonometric parallaxes $p'' = p''(\mathrm{tr})$ and the formula* $M_\lambda = m_\lambda^\circ + 5 + 5 \lg p''$ $\left(p'' = 206\,265 \frac{\text{a.u. (km)}}{r\,(\text{km})} = \frac{1}{r\,(\text{pc})} \right)$. Unfortunately, the trigonometric parallaxes measured with meridian transit instruments are highly unreliable (with errors of up to 0".2). Exact parallaxes are determined with a micrometer and a heliometer. At present the trigonometric parallaxes are derived almost entirely by photographic methods (using long-focus astrographs). Even in this case the accuracy is not particularly high, however, and only seldom permits measuring parallaxes of less than 0".01. Indeed, for a focal distance of 10 m, this angular size corresponds to a mere 0.49 micron in the focal plane of the astrograph, while the apparent diameter of the star on a photographic plate is as large as 40—200 microns! We can thus measure only the p'' of stars with distances not greater than r (pc) $= \frac{1}{p''} = \frac{1}{0''.01} = 100$ pc. At present there are some 160 stars only with known trigonometric parallaxes $p'' > 0''.1$ and about 3900 stars with $p'' > 0''.01$. It is obvious that both the large and the small parallaxes have the same absolute errors, and the relative accuracy is thus different.

* Corrected for galactic absorption, $m_\lambda = m_\lambda^\circ + A(\lambda)$.

The trigonometric parallaxes of some stars were measured photographically at numerous observatories and are known on the average with high accuracy. For example, for Barnard's star, the data of 11 observatories give $p'' = 0''.546 \pm 0''.003$.

Up to 1952, Jenkins's Yale Catalogue of trigonometric parallaxes listed 5822 stars for which the parallax had been measured at least once. Some of them show extremely small parallaxes (about $0''.001$), which are in all probability unreliable. Moreover, the internal accuracy of the parallaxes and the biased errors which emerge when the parallaxes measured at different observatories are compared should be kept in mind. According to the last IAU report (Moscow, 1958), there are currently 10,823 individual measurements of trigonometric parallaxes.

The determination of parallaxes smaller than $0''.01$, i.e., the parallaxes of far stars, requires a larger base than one astronomical unit ($A = 1$ a.u. $= 1.495 \cdot 10^8$ km). The annual path of the Sun can be used as the new base:

$$h_\odot = V_\odot \left(\frac{\text{a.u.}}{\text{year}}\right) \cdot 1\,(\text{year}) = V_\odot \left(\frac{\text{km}}{\text{sec}}\right) \frac{n(\text{sec})}{A\,(\text{km})} =$$

$$= V_\odot \left(\frac{\text{km}}{\text{sec}}\right) \frac{31\,557\,000\ \text{sec}}{1.495 \cdot 10^8\ \text{km}} = V_\odot \left(\frac{\text{km}}{\text{sec}}\right) \cdot \frac{1}{4.74} \approx \frac{19.5}{4.74} = 4.1\ (\text{a.u.})$$

If λ is the angular distance to the apex and v is the parallactic angular displacement of the star over the sky, we have as usually

$$\frac{h_\odot}{r} = \frac{v\,(\text{rad})}{\sin\lambda} = \frac{v''\sin 1''}{\sin\lambda}; \qquad \left(\frac{\bar{h}_\odot}{r}\right)'' = \frac{\bar{v}''}{\sin\lambda} = \bar{p}_s''. \tag{86}$$

The stellar parallax $\bar{p}_s''$ is called secular parallax. Hence for the ordinary parallax we get

$$p_v\,(\text{rad}) = \frac{v\,(\text{rad})}{h_\odot \sin\lambda} = \frac{v''\sin 1''}{h_\odot \sin\lambda}, \tag{87}$$

$$\bar{p}_v'' = \frac{v''}{h_\odot \sin\lambda} = \frac{v''}{V_\odot \left(\frac{\text{a.u.}}{\text{year}}\right) \cdot 1(\text{year}) \sin\lambda} = \frac{v'' A}{V_\odot \left(\frac{\text{km}}{\text{sec}}\right) n \sin\lambda} =$$

$$= 4.74 \frac{v''}{V_\odot \left(\frac{\text{km}}{\text{sec}}\right) \sin\lambda} = 4.74 \frac{\bar{p}_s''}{V_\odot} \qquad \left(\frac{A}{n} = \frac{1.495 \cdot 10^8}{3.156 \cdot 10^7} = 4.74\right). \tag{88}$$

To eliminate the peculiar motions of stars, one computes the mean parallax $\bar{p}_v$ of a group of stars. In this case, we obtain by the least squares method

$$\bar{p}_v'' = 4.74 \frac{v''\overline{\sin\lambda}}{V_\odot \overline{\sin^2\lambda}} = \frac{4.74}{V_\odot} \bar{p}_s''. \tag{89}$$

Similarly from the peculiar component of the proper motions $\bar{\tau}$ we obtain (taking the absolute values)

$$\bar{p}_\tau = 4.74 \frac{|\bar{\tau}|}{|\bar{V}_p|}. \tag{90}$$

Here $|\bar{V}_p|$ is the mean peculiar velocity of the star in km/sec. The parallaxes derived in this way can naturally be applied only to determine the mean absolute magnitudes of a large group of stars.

2. Group parallaxes associated with the common motion of stars can also be used. To this end one uses moving clusters and the effect of galactic rotation. In the former case the mean parallax is

$$\bar{p} = \frac{4{,}74}{N\bar{V}} \sum_{i=1}^{N} \frac{\mu_i}{\sin \Lambda_i}. \tag{91}$$

Here Λ_i is the angular distance of the individual star from the cluster radiant, N is the number of stars, $\bar{V}$ is the velocity of the cluster, and μ_i are the proper motions of the stars.

In the second case the mean distance and hence the parallax are determined from the radial velocity V_r corrected for the proper motion of the Sun:

$$V_r = \bar{Ar} \sin 2L \cos^2 b = Ar \sin 2(l - l_0) \cos^2 b. \tag{92}$$

The most acceptable value of Oort's constant is apparently $A = 17.2$ km/sec/kpc (according to Mel'nikov (1944), Petrie (1956), and others). Other authors give $10 \leqslant A \leqslant 25$ (km/sec)/kpc and occasionally the values even range between wider limits. Group parallaxes of the Pleiades, Perseus, Orion, and Scorpius — Centaurus clusters give for the absolute magnitudes of the stars in these moving clusters $-0^m.1$, $0^m.0$, $-2^m.0$, and $-0^m.5$ (for 8, 20, 17, and 11 stars, respectively). Spectroscopic methods give $+0^m.1$, $+0^m.5$, $-2^m.0$, and $-1^m.0$, respectively. Hence it follows that group parallaxes are on the whole not less accurate than the spectroscopic parallaxes (especially for the near clusters).

3. The dynamic parallaxes of binary stars can also be used. In this case, by Kepler's third law, setting P for the period in siderial years, we get

$$p'' = \frac{a''}{\sqrt[3]{\left(\frac{\mathfrak{M}_1}{\mathfrak{M}_\odot} + \frac{\mathfrak{M}_2}{\mathfrak{M}_\odot}\right) P^2}}. \tag{93}$$

Visual binaries are often observed which have common proper motion but the relative motion of the components is too small for orbit determination. For very large revolution periods, only part of the orbit can be observed. In this case one uses the formula obtained independently by Russell and Hertzsprung (1911), which can be written in the form

$$p'' = \sqrt[3]{\frac{s''(\omega'')^2}{4\pi^2 \left(\frac{\mathfrak{M}_1}{\mathfrak{M}_\odot} + \frac{\mathfrak{M}_2}{\mathfrak{M}_\odot}\right) K}}. \tag{94}$$

Here s'' is the apparent angular distance of the components in seconds of arc, $s'' = rp'' \sin i$ (r is the distance between the components in a.u.), ω'' is the apparent relative displacements of the components in angular seconds per year (this is the angular velocity equal to $vp'' \sin j$, where v is the linear velocity in A. U. per year), and $K = \left(2 - \frac{r}{a}\right) \sin i \sin^2 j = \left(2 - \frac{s''}{ap'' \sin i}\right) \sin i \sin^2 j$, where i is the angle between the radius-vector and the line of sight, j is the angle between the velocity vector and the line of sight. Since the variance of masses is small compared to the variance of other quantities, we may

take $\frac{\mathfrak{M}_1}{\mathfrak{M}_\odot}+\frac{\mathfrak{M}_2}{\mathfrak{M}_\odot}=2$; the constant K for most individual stars is not known, but its probable mean value for binaries can be computed. It is assumed in these calculations that different inclinations of the binary orbits are all equiprobable and that the rms eccentricity is $\bar{e}=0.561$. Then $K=0.429$. The above equation was obtained by making use of Kepler's equation with orbital linear velocity v. If we now use the parabolic velocity v_{par}, which is always greater than the orbital velocity, and take, as is usually done, $a=\infty$, we can calculate the lower limit value* for the binary parallax: $p''_{min}=\sqrt[3]{\frac{s''(\omega'')^2}{4\pi^2\cdot 2\cdot 2}}=\sqrt[3]{\frac{s''(\omega'')^2}{16\pi^2}}=0.185\sqrt[3]{s''(\omega'')^2}$. If the components of a visual binary can be subjected to spectrography, which is generally feasible for angular distances of not less than 1'', the semiaxis a (km) can be found from the radial velocity, and since a'' is known, the parallax is easily calculated. In This way Belopol'skii obtained back in 1898 the p'' for γ Vir and γ Leo ($\frac{4''.02}{102}=0''.039$ and $\frac{2''.00}{98.1}=0''.020$, respectively).

4. For eclipsing binaries the relative radii $\frac{R}{a}$ of the components can be found from the photometric light curve.

Expressing the stellar masses in units of solar mass ($\mathfrak{M}_\odot$), the period P in siderial years, and a in astronomical units, we further have

$$\mathfrak{M}_1+\mathfrak{M}_2=\frac{a^3}{P^2} \tag{95}$$

and from the mass — luminosity relation (see below) and others

$$L\sim\mathfrak{M}^{10/3}.$$

Further

$$\left.\begin{aligned} L&\sim R^2T_e^4,\\ R&\sim a. \end{aligned}\right\} \tag{96}$$

L_v can now be found and hence M_v. Using the bolometric correction and the relation $M_b=m_b''+5+5\lg p''$, we thus obtain the parallax p''. All this of course implies knowledge of the masses $\mathfrak{M}_1$ and $\mathfrak{M}_2$.

5. Various indirect methods are available for estimating the distances of stars from the Sun (their parallaxes): the period — luminosity relation for the cepheids and other characteristics of variables, in particular, the maximum brightness of novae; the intensities of Ca II, Na I and other interstellar lines (O. Struve, 1929, and others), observations of the 21 cm radio lines (emission and absorption), etc. Radio measurements give the distances to the spiral arms of the Galaxy where the hot supergiants are found (O and B type stars). If the distances are known, the mean absolute magnitudes of the hot supergiants can be computed; this is a unique achievement as no other method will work for these stars, especially those of type OIa.

Various other formulas are available for the calculation of mean absolute magnitudes and hence mean parallaxes from observation data. Thus, for example, for stars with known and noticeable proper motions $M_v=a+bH+cH^2\approx a+bH=-1.675+0.645H$. Here H is the "reduced proper motion," $H=m+5+5\lg\mu$, where μ is the proper motion.

* In particular, when the orbital plane is perpendicular to the line of sight.

6. The most extensive observation series of parallax determination were carried out with spectrographs. These are the so-called spectroscopic parallaxes. More precisely, this method determines not the parallax but the absolute magnitude, which in conjunction with the apparent magnitude gives the parallax. The correlation between the intensity of the line $\lambda = 4077$ Å and the M_v of stars was first noted by Hertzsprung back in 1906.

Since we use the relation

$$M_\lambda = m_\lambda^\circ + 5 + 5 \lg p'' = m_\lambda - A(\lambda) + 5 + 5 \lg p'', \tag{97}$$

the parallax error is (for $|dM| = \sigma_M$, etc.)

$$2.17^2 \frac{\sigma_p^2}{p^2} = \sigma_M^2 + \sigma_{(m)}^2 \tag{98}$$

($dm_\lambda \ll dM_\lambda$) so that the relative error is virtually the same for small and large parallaxes. If we assume that the exact m_λ is known, we have

$$dM = 2.17 \frac{dr}{r}, \quad dp \approx 0.5 p\, dM.$$

For high-dispersion instruments (slit spectra) $dM \approx 0^m.3$ and therefore $\frac{dp}{p} \approx 0.14 = 14\%$, whereas for low-dispersion instruments (slitless spectra) $dM \approx 1^m.0$ and $\frac{dp}{p} \approx 0.46 = 46\%$. For practical determination of M_λ from known parallaxes one plots calibration curves which relate the absolute magnitudes M_λ to the line intensities W_λ.

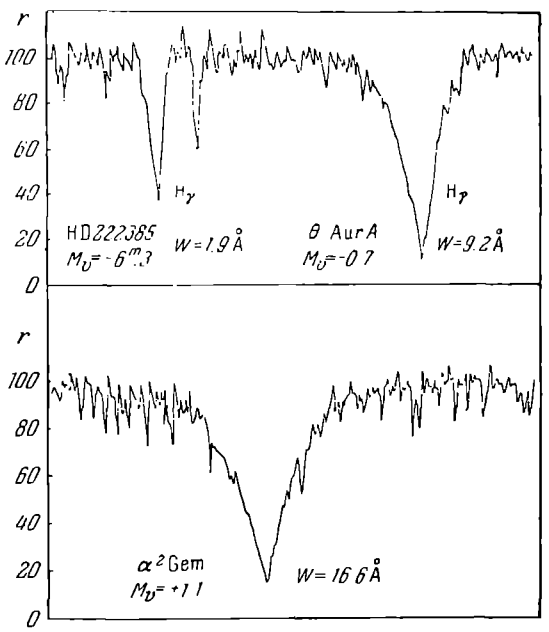


FIGURE 7

For the early-type B—A stars, photometric measurements give W_λ for the hydrogen lines, H_γ say. Figure 7 shows the H_γ profiles from the spectra of several A-type stars: HD 222385, θ Aur A, and α^2 Gem with absolute magnitudes $-6^m.3$, $-0^m.7$, $+1^m.1$ (luminosity classes Ia—Ib, IV—V, V) and equivalent line widths of 1.9, 9.2, and 16.6 Å, respectively. The correlation with absolute magnitude is clearly obvious.

Petrie and Maunsell (1949) obtained the equivalent widths $W(H_\gamma)$ for various absolute visual magnitudes M of 169 Sp = B8—A3 stars. Their results are listed in Table 8.

TABLE 8

$W(H_\gamma)$, Å	B8—A3 stars						
	2.71	7.53	10.18	12.50	14.17	16.10	18.24
M_v	−5.73	−1.16	+0.25	+0.82	+1.06	+1.25	+1.41

The average dependence of M_v on $W(H_\gamma)$ for these 169 late B and early A type stars (Table 8) can be expressed by the relation

$$M_v = 1.69 - 13.25 \cdot 10^{-0.095 W(H_\gamma)}. \qquad (99)$$

Similarly, Mel'nikov (1954) obtained for B8—A7 stars

$$\lg W\left(\frac{H_\gamma + H_\delta}{2}\right) = 1.025 - 0.115\, M_v. \qquad (100)$$

Fair results are obtained for the absolute magnitudes of B—A stars from the maximum number of lines in the Balmer series or from the number of the last line n_m (the quantum number of the uppermost level), corrected for instrumental and other distortions. Thus, Miczaika (1948) obtained for B0—B3 stars

$$M_v = -0^m.56\,(n_m - 10) + 1^m.21. \qquad (101)$$

Mel'nikov obtained for A stars (Table 9)

$$\overline{\lg n_m} = 1.310 - 0.025 M_v. \qquad (102)$$

The above methods are clearly applicable only if the star spectra show the hydrogen Balmer lines.

TABLE 9

$\overline{\lg n_m}$	1.477	1.459	1.384	1.310	1.310	1.277	1.999	1.279
$\overline{M}_v$	−5.40	−3.60	−2.57	−1.03	−0.34	+0.64	+1.56	+2.30
Number of stars	3	1	3	3	20	25	17	2

A new and fundamentally different method for the determination of luminosities (and spectral types) of B—A stars and others from continuous spectra was developed by Chalonge and Barbier and co-workers (1950—1959). They used the Balmer discontinuity D_{λ_0} near the limit of the Balmer series $\lambda_0 \approx 3646$ Å:

$$D_{\lambda_0} = \lg \frac{I_{\lambda_0+\varepsilon}}{I_{\lambda_0-\varepsilon}} \qquad (103)$$

and the wavelength λ_1 corresponding to the extrapolated midpoint of the discontinuity as their parameters. A third parameter was added to this two-parameter classification — the spectrophotometric violet gradient φ_b.

The graph illustrating the results of the French astronomers is shown in Figure 8.

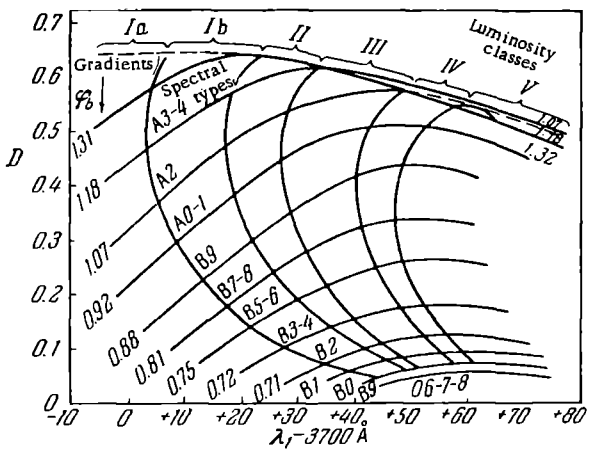


FIGURE 8

The most accurate determinations of the spectroscopic absolute magnitudes of stars from O to M are based on line intensity ratios. The intensity ratios or differences were chosen for the determination of spectral types and luminosity classes when this method, soon after the original publication by W. Adams and A. Kohlschutter (1914), was being rapidly developed at the Mount Wilson Observatory (Adams, Joy, et al.). The lines used in the determination of the spectrum were those sensitive to temperatures changes, and the lines for the determination of M_v were sensitive to surface acceleration. Eventually some 15 observatories joined the project and the number of spectroscopic parallaxes is now over 30,000. The total number of trigonometric parallaxes (sufficiently reliable) is about 8000, and the parallaxes determined by other methods number about 4000 (Parenago's catalogue, 1947).

As an illustration, Figure 9a plots the intensity ratio of the 4144 Å helium line to the H_δ line for B-type stars (Lockyer Observatory). Figure 9b gives the same plot for K8 and F8 types (Mount Wilson). The

intensity ratios (D) of the line pairs 4215 Å/4250 Å, 4345 Å/4415 Å, 4408 Å/4415 Å, 4456 Å/4462 Å, 4456 Å/4495 Å (Mount Wilson) make it possible to calculate the absolute magnitude from the equalities

$$M_v = 5.6 - 1.6D \ (\text{F8} - \text{G6}), \quad M_v = 6.8 - 1.8\overline{D} \ (\text{G6} - \text{K9}).$$

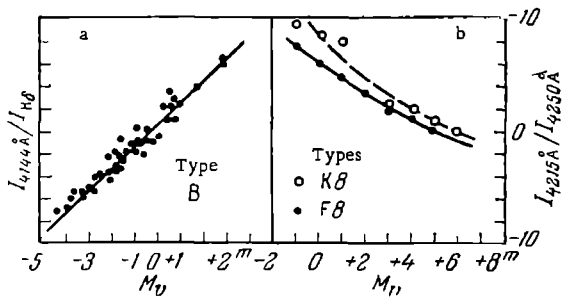


FIGURE 9

The line pairs for various spectral types are listed in Table 10 (Mount Wilson).

TABLE 10

Luminosity criterion, Å	Comparison line, Å	Spectral types							
		A	F	dG	dK	dM	gG	gK	gM
4007 SrII	4071	+	−	+	+		+	−	+
4161 TiII	4167	+	+	+			−		
4196 FeI	4190			+	+			+	
4207 FeI	4202								+
4215 SrII	4250	+	+	+	+		−	−	
4233 FeII	4236	+	+				−	−	
4246 ScII	4250	+	+	+			−		
4258 FeI, FeII	4260			+	+				+
4290 CrI, TiII	4271	+	+						
4318 CaI	4321			+	+	+			
4324 CrI	4321		−	+			+	+	
4340 H_γ	4325							−	+
4375 ScII, YII	4404	+	+				+		
4379 VI	4376				+	+			
4399 TiII	4404	+	+	+			+		
4408 FeI, TiII	4465			+	+		−	+	
4435 CaI	4415				−	+			
4454 CaI	4461		+	+		−			
4489 FeI, FeII	4494							+	+
4535 TiI	4415					+	−		
4586 CaI	4572						+		
4607 SrI	4603						−		
4861 H_β	4871							+	+

Table 11 shows the variation of intensity ratios for spectral types F0 through G0 (J. Hynek, 1935). In their determinations of M_v, Adams and co-workers (1922) came to the conclusion that the spectral types of stars should be divided into groups according to the character of the lines. They distinguished between stars with sharp and diffuse spectral lines. Accordingly

they added a lower-case s to the spectral type symbols of the first category and n to those of the second category. The group c according to Maury (1897) was also retained. The stars c and even s have higher absolute brightness than n. The difference in absolute magnitude between s and n stars in subclass B0 is not large, but it is very pronounced in classes B2—F2, where it is equal to a fraction of stellar magnitude. Later additional qualifying letters were added to the spectral classes, e.g., m for A stars (now designated ML) with enhanced metallic lines according to Weaver (1952), sd (or VI) for subdwarfs, wd for white dwarfs (alternatively D, VIIa or VIIb), and so on.

TABLE 11

Intensity ratio	Absolute magnitude, M_v	Spectral types			
		F0	F5	F8	G0
$\frac{\text{FeI 4071 Å}}{\text{SrII 4077 Å}}$	$-2^m.0$	0.56	0.70	0.61	0.81
	$+3\ .5$	0.97	1.17	1.32	1.54
$\frac{\text{TiII, ZrII 4161 Å}}{\text{MgI 4167 Å}}$	$-2\ .0$	1.70	1.80	2.00	1.00
	$+3\ .5$	0.63	0.79	0.69	0.45

As the spectral classification became more developed and the absolute magnitudes of the early-type B—A stars improved (1923—1927) more complex combinations of indices were introduced for different varieties of lines. Thus, at the Lockyer Observatory, the standard spectral type symbols were qualified by lower-case letters or combinations, ss, s, ns, n, nn, in the order of decreasing luminosities.

Note that in the Mount Wilson catalogue of spectroscopic absolute magnitudes of 4179 stars the probable error is $\sigma_{M_v} = \pm 0^m.27$. If we take $\sigma_{m_v} = 0^m.1$, then from (98) $\frac{\sigma_p}{p} = 0.13 = 13\%$ irrespective of p.

Note that the spectroscopic method of determination of absolute magnitudes was discovered by observers and only at a later stage, after the development of ionization theory, was it explained theoretically.

At present, as we have noted, there are over 30,000 of spectroscopic stellar parallaxes determined with various high- and low-dispersion slit spectrographs and prism cameras. For $p'' > 0''.05$ (i.e., $r < 20$ pc), the spectroscopic data are less accurate than the trigonometric method. For far stars, however, the spectroscopic parallaxes are much more accurate. According to one of the first studies of R. V. Kunitskii in this direction, the "weight" of the spectroscopic parallaxes for $p'' = 0''.30$, $0''.10$, and $0''.02$ (in units of the "weight" of the spectroscopic parallax for $0''.015$) is 0.05, 0.4, and 10.0, respectively. For stars with anomalous spectra, however, p''(sp) is worse than p''(tr) (e.g., for Arcturus p''(tr) = $0''.08$, whereas p''(sp) = $0''.1—0''.2$).

The top part of Table 12 gives a comparison of the spectroscopic p''(sp) and the trigonometric p''(tr) parallaxes for a number of bright stars. We see that the fit is satisfactory. For weak stars (lower part of the table) the fit is considerably poorer (the average difference p''(sp) — p''(tr)

TABLE 12

Star	α Aur	α Ori		α CMa		α CMi		α Cen		β Gem
p''(sp) p''(tr)	0".085 0".073	0".012 0".005		0".40 0".375		0".34 0".288		0".78 0".751		0".11 • 0".093 ••
M_v ••• / Sp	−3.0 − +1.9 p''(sp)−p''(tr)	Number, N	+2.0 − +4.9 p''(sp)−p''(tr)	Number, N	+5.0 − +7.9 p''(sp)−p''(tr)	Number, N	+8.0 − +13.3 p''(sp)−p''(tr)	Number, N	Total p''(sp)−p''(tr)	Number, N
F0—F8	+0".003	12	+0".004	62	+0".018	14	—	—	+0".006	88
F9—G7	+0 .005	22	+0 .007	35	+0 .007	34	—	—	+0 .007	91
G8—K2	−0 .004	38	−0 .011	21	+0 .002	35	−0".021	2	−0 .004	96
K3—K9	+0 .011	16	−0 .108	1	+0 .014	23	+0 .005	17	+0 .008	57
Ma—Md	+0 .004	14	−0 .049	1	—	—	+0 .008	13	+0 .004	28
Average	+0".002	102	+0".001	120	+0".008	106	+0".004	32	+0".0037	360

• From Gen. Cat. of Star Parallaxes (F. Schlesinger, L. E. Jenkins), Yale Univ. Obs., 1935.

•• From Gen. Cat. of Trigonometric Parallaxes (L. E. Jenkins), Yale Univ. Obs., 1952.

••• From 360 p''(tr) and 500 p''(sp): ApJ 46, 316, 1917; Hd. d. Ap., V/I, 457, 1932. (Average difference p''(sp) − p''(tr) = = 0".026 ignoring the sign.)

for 360 stars, ignoring the sign, is 0".026, since interstellar absorption distorts the apparent magnitude m_t used in parallax calculation). On the whole, for the far stars, spectroscopic parallaxes are nevertheless more accurate than trigonometric: for example, for a small spectroscopic parallax 0".0030 the accuracy is ±0".0004, which is quite satisfactory. Moreover, spectroscopic parallaxes are never negative, unlike some trigonometric paraliaxes. Originally the spectroscopic absolute magnitudes were determined from slit spectrograms (reciprocal dispersion of 16 and 36 Å/mm for H_γ), but at a later stage spectra obtained with prism cameras began to be used (reciprocal dispersion 100—500 Å/mm). These measurements were conducted on a particularly large scale in Harvard, Uppsala, Stockholm, Potsdam, Bergedorf, Cleveland, Abastumani, Crimea, etc.

The photoelectric methods of narrow-band (50—100 Å) integrated-light photoelectrocolorimetry began to be successfully applied to the determination of spectral types and luminosities with the aid of prism cameras and interference filters. These investigations were the outgrowth of Lindblad's method, first proposed 30 years ago. Reflectors with photoelectric photometers are now used for this purpose.

The intensity ratio in two spectral regions is measured (λ4240/λ4170); it is used as a measure of the intensity of the CN band and is fairly sensitive to the luminosity class. Conversely, the ratios λ4220/λ4240 and λ3910/λ4030, which are a measure of the intensity of the band and the K line of Ca II are largely sensitive to temperature. The final McDonald and Copenhagen classification of F—G stars (photoelectric observations) use the indices $I(H_\beta)$ and $I(H_D)$, i.e., the combined intensity near the H_β line and near the Balmer discontinuity $D_{\lambda 0}$. The luminosity classes of M_v are defined in terms of the quantity

$$I(H_\beta) = -2.5\left\{\frac{1}{2}[\lg I(4700\text{Å}) + \lg I(5000\text{Å})] - \lg I(4861\text{Å})\right\} + \\ + \text{const} = \frac{1}{2}(m_{4700} + m_{5000}) - m_{4861}, \qquad (104)$$

which decreases from class Ia to class V. The spectral types are determined in terms of the quantity

$$I(H_D) = -2.5\{[\lg I(4030\text{Å}) - \lg I(3550\text{Å})] + [\lg I(4030\text{Å}) - \lg I(4500\text{Å})]\} + \\ + \text{const} = (m_{4030} - m_{3550}) + (m_{4030} - m_{4500}), \qquad (105)$$

which decreases from F to G. The extreme wavelengths are so chosen that the effect of interstellar absorption is automatically eliminated; this can be done only approximately, of course, using the λ^{-1} law. When the two indices have been determined, the luminosity classes and the spectral types are read off the index vs. absolute magnitude curve (for bright stars).

Morgan, Meinel, and Johnson (1954) used a prism camera with exceedingly low dispersion (30,000 Å/mm) for the determination of the luminosity classes and spectral types. The red stars give "toadpole" images on the photographic plate, and the images of white-blue stars are "lenticular." Even with this low dispersion, the spectral types and the luminosity classes could be determined by an experienced observer from a simple examination of the spectrum. This method was applied in the USSR, though in a somewhat modified form, at Pulkovo by Mel'nikov and Stoyanova (1958).

It should be noted that the distances of stars and hence their parallaxes and absolute magnitudes can be obtained by an indirect method — from the intensities (equivalent widths) of the interstellar calcium $W(K)$ and sodium $W(D)$ lines. If the equivalent widths are expressed in km/sec $\left(v = c\frac{\Delta\lambda}{\lambda} = c\frac{W_i}{\lambda_i}\right)$ and the distance in parsecs, then according to Beals and Oke (1954) $r = 34.83\ W(K)$ and $r = 30.75\ W(D)$ (from 1820 B-type stars).

In 1943 Morgan, Keenan, and Kellman, while working on the Yerkes Spectral Classification MKK (see §7), identified the luminosity class by Roman numerals following the spectral type symbol. Later, in 1953, Morgan and Johnson improved the MKK classification and linked it to the color index scales $(B - V)$ and $(U - B)$. This classification is designated MK (see §7). The relation of the Yerkes (MK) classification to the classical Harvard (HD) classification is shown in Table 13 (see also Table 17). This subject is treated in more detail in the next section.

TABLE 13

Spectral type MK	O6	O8	O9	B0	B1	B2	B3	B5	B8	A0	A2	A3	A5	A7	F0
Spectral type HD	O6.6	O8.2	O9.2	B0.1	B1.0	B2.3	B3.0	B4.9	B7.8	A0.0	A1.6	A2.3	A4.8	A6.4	A9.7
Spectral type MK	F2		F5	F8	G0	G2		G5	K0	K3		K5	K8		M0
Spectral type HD	F1.5		F4.9	F8.4	F9.6	G1.2		G4.4	K0	K1.8		K5	K8		M0.5

The luminosity class in MKK and MK classifications is designated by the Roman numerals Ia, Ib, II, III, IV, V, VI, VII for most luminous (Ia) and less luminous (Ib) supergiants, bright and normal giants (II and III), subgiants (IV), main-sequence stars (V), subdwarfs and white dwarfs (VI and VII).* All this is discussed in more detail in Chapter 18.

The relation between the spectrum (Sp), the luminosity (M_v), and the color index $(B - V,\ U - B)$ for luminosity classes V and III is given in Table 14.

Figure 10 schematically shows the Hertzsprung—Russell diagram (abbreviated as HR diagram) in the usual coordinates M_v vs. Sp (MK) but using the new spectral classification MK. The dashed vertical lines join points of equal temperature. The luminosity classes are marked on the right (from Ia to V); they determine the shape of the given sequence of spectral classes for fixed average luminosity. In Figures 11 and 12 the HR diagram is plotted in new coordinates, M_v vs. $(B-V)$ and M_v vs. $(U - B)$ from the data of Morgan and Johnson (1952—1953). The color indices $(B - V)$ and $(U - B)$ are determined photoelectrically to very high precision, which is not so for the spectral types of the HR diagram in the M_v vs. Sp coordinates. Naturally the colors $(B - V)$ and $(U - B)$ show good correlation

* These two symbols are not commonly used.

TABLE 14

Luminosity classes V and III							
Sp	M_v	$B-V$	$U-B$	Sp	M_v	$B-V$	$U-B$
B0	$-4^m.3$	$(-0^m.32)$	$(-1^m.13)$	G5V	$+5^m.2$	$+0^m.68$	$+0^m.21$
B1	−3 .2	−0 .28	−1 .00	G8	+5 .5	+0 .70	+0 .24
B2	−2 .4	−0 .24	−0 .86	K0	+5 .8	+0 .82	+0 .48
B3	−1 .7	−0 .20	−0 .71	K1	+6 .0	+0 .86	+0 .54
B5	−0 .7	−0 .16	−0 .56	K3	+6 .5	+1 .01	+0 .89
B7	+0 .1	−0 .13	−0 .47	K5	+7 .2	+1 .18	+1 .12
B8	+0 .5	−0 .09	−0 .29	K7	+7 .7	+1 .37	+1 .26
B9	+0 .8	−0 .05	−0 .16	M1	+9 .2	+1 .48	+1 .21
A0	+1 .1	0 .00	0 .00	M3	+10 .0	+1 .49	+1 .10
A1	+1 .4	+0 .05	+0 .05	M5V	+12 .9	+1 .69	+1 .24
A3	+1 .7	+0 .09	+0 .07	—	—	—	—
A5	+2 .0	+0 .15	+0 .09	—	—	—	—
A7	+2 .3	+0 .19	+0 .08	—	—	—	—
F0	+2 .7	+0 .30	+0 .02	—	—	—	—
F2	+3 .0	+0 .37	0 .00	C8III	+0 .5	+0 .95	+0 .72
F5	+3 .5	+0 .44	0 .00	K0	+0 .4	+1 .01	+0 .86
F6	+3 .7	+0 .47	−0 .02	K2	0 .0	+1 .16	+1 .20
F8	+4 .1	+0 .53	+0 .02	K3	−0 .5	+1 .30	+1 .44
G0	+4 .4	−0 .60	+0 .06	K5	−1 .1	+1 .52	+1 .84
G2V	+4 .7	+0 .64	+0 .16	M2—3	−1 .2	+1 .57	+1 .86

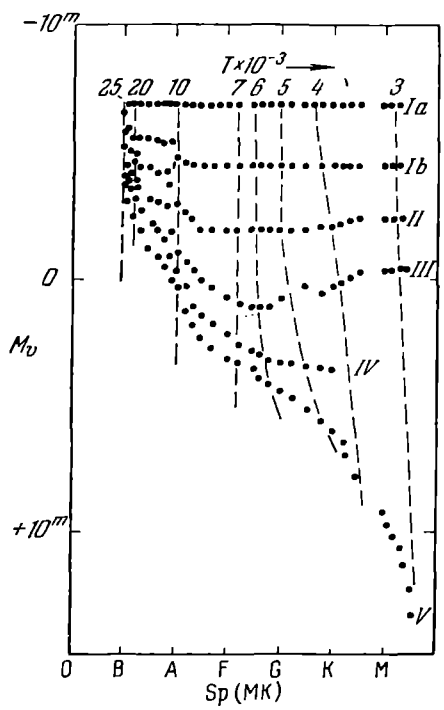


FIGURE 10

with temperature. The curve of the $(B-V)$ and $(U-B)$ color indices, its nonlinearity and nonmonotonic variation are responsible for the considerable difference in the appearance of the HR diagrams of main sequence stars when the color $(B-V)$ or even $(U-B)$ is laid off the abscissa.

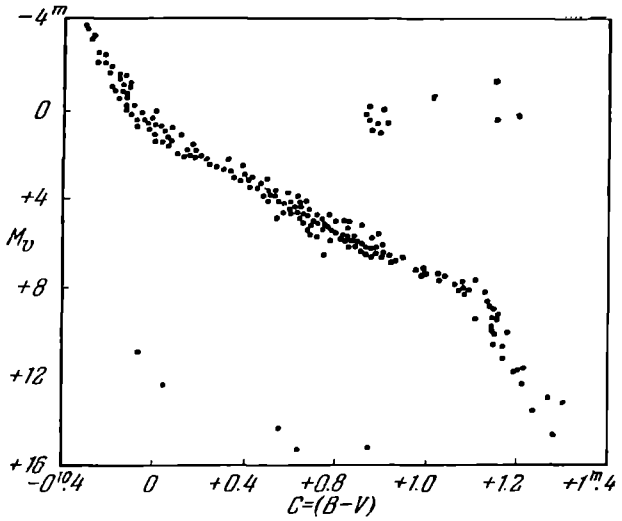


FIGURE 11

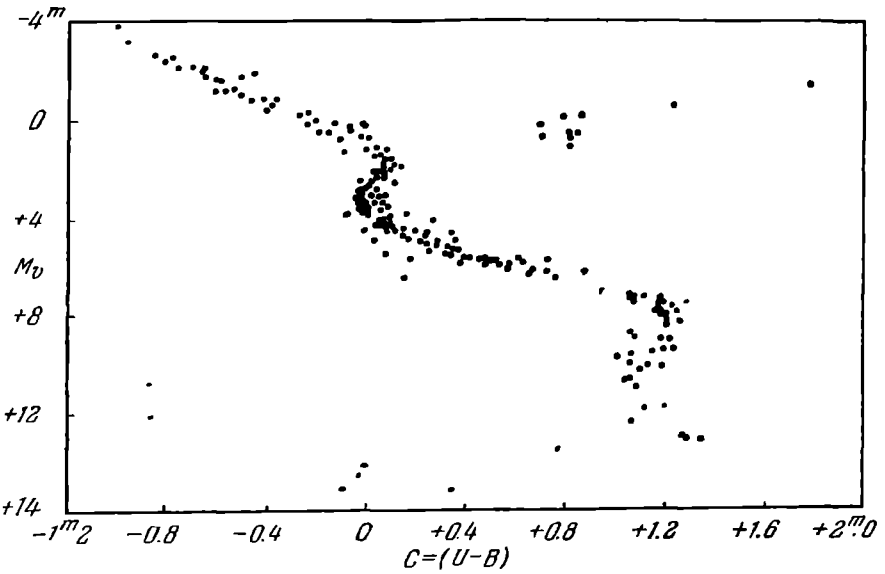


FIGURE 12

The $(B-V)$color and the spectral types of stars of constant luminosity are conversely related by a monotonic dependence, and they are therefore used more often than the $(U-B)$colors. Unfortunately, the $(B-V)$color, like any other color, is affected by interstellar extinction. However, Morgan and Johnson managed to overcome this difficulty. Let $(B-V)_0$, $(U-B)_0$ be the true colors and $(B-V)$, $(U-B)$ the observed colors. Let further E_{B-V} and E_{U-B} be the corresponding color excesses. We can thus write

$$\left.\begin{aligned}(B-V)&=(B-V)_0+E_{B-V},\\(U-B)&=(U-B)_0+E_{U-B}.\end{aligned}\right\} \tag{106}$$

According to Morgan and Johnson, for O—B stars

$$\frac{E_{U-B}}{E_{B-V}}=\text{const}=0.72\pm0.03.$$

We define a new quantity

$$Q=(U-B)-\frac{\overline{E_{U-B}}}{\overline{E_{B-V}}}(B-V)=(U-B)_0-\frac{\overline{E_{U-B}}}{\overline{E_{B-V}}}(B-V)_0. \tag{107}$$

If the ratio $\frac{E_{U-B}}{E_{B-V}}$ is independent of interstellar absorption, Q is also independent of this factor. For different spectral types (O5, B0, B5, A0), Q takes on different values: −0.93, −0.90, −0.44, and 0.00. Hence, Q can be used as a criterion of the spectral type from stellar continuum parameters measured with fairly wide-band filters coupled to a photomultiplier.

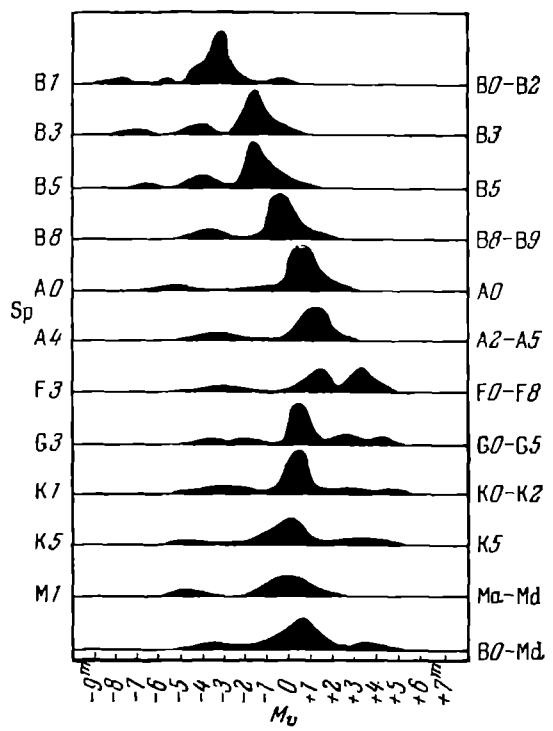


FIGURE 13

The true color can thus be related to the measured Q (Sp) for various spectral types:

$$(B-V)_0 = -0.09 + 0.337Q\ (\mathrm{Sp}). \tag{108}$$

Hence, for the color excess

$$E_{B-V} = (B-V) - (B-V)_0 = (B-V) - 0.337Q + 0.009. \tag{109}$$

The observed colors of O6—A0 stars can thus be used to investigate the color excess, i.e., the interstellar absorption.†

As a supplement to our discussion of absolute stellar magnitudes, Figure 13 shows the dispersion of absolute magnitudes of various spectral types (in the HD system) up to apparent magnitude of 6^m.

§7. Spectral classification of cold stars. Criteria of the MK classification. The HD classification, described in Vol. I of this Course, includes the spectral types of cold stars M, S and R, N. The first type is characterized by the presence of TiO bands, the second by the presence of ZrO bands, and the last two by bands of carbon compounds CH, C_2, C_3, CN, etc. By now the two types R, N have been combined into a single temperature class C subdivided into C0 to C9. It is characterized by the presence of the following molecular bands: 1) C_2 (Swan bands) $A^3\Pi_g - X^3\Pi_u$ (0.2), (0.1), (0.0), (1.0) and (2.0) in the regions 6191, 5636, 5165, 4737, 4383 Å, which respectively include 6, 5, 3, 4, and 3 main bands; 2) CN $B^2\Sigma - X^2\Sigma$ of the sequences −2, 1, 0, 1 with heads at 4606, 4216, 3883 and 3590 Å, etc. The old Harvard HD classes R0, R3, and R5 roughly correspond to the new C1, C2, C4. No such correspondence is observed for the old N type, however. For example, the extremely cold star WZ Cas (whose spectrum shows the 6708 Å Li I line!) was classified as N1 (HD), whereas in the new classification it is treated as C9 (MK). According to Keenan and Morgan (1951), the effective temperatures T_e of the C subclasses range from 4500° (C0) to 3450° (C5), as we see from Table 15.

TABLE 15

T_v *	T_e	C type **	HD type (same T_e)
4100	4500° K	C0	G4—G6
—	4300	C1	G7—G8
3500	4100	C2	G9—K0
—	3900	C3	K1—K2
2900	3650	C4	K3—K4
—	3450	C5	K5—M0
2300	—	C6	M1—M2
—	—	C7	M3—M4
2100	—	C8	—

* R. Bouigue.—Ann. d'Astrophys. 17, 35, 1954.

** W.W. Morgan, P.C. Keenan.—ApJ 94, 501, 1941.

† A certain difficulty is encountered due to the progressive growth of the effective wavelengths of the colorimetric system for progressively farther stars, which are more susceptible to the distorting action of interstellar absorption (Nikonov, 1953, and others).

The same table lists the temperatures T_v derived from the Cn vibration bands (Bouigue, 1954) and the spectral types G—M from the temperatures T_e which correspond to the classes C0—C7. Note that half the sum of the intensities of the D_1 and D_2 Na I lines shows good correlation with the C subclasses: it monotonically increases from C0 to C9.

TABLE 16*

Criteria	
spectral type	luminosity class
O5, O9: $\frac{\text{He II } 4541}{\text{He I } 4471}$	$\frac{\text{C III } 4068}{\text{Si IV } 4089}$, $\frac{\text{Si IV+He I } 4119}{\text{He I } 4144}$, $\frac{\text{He I } 4387}{\text{He II } 4541}$
B0, B0.5, B1: $\frac{\text{Si III } 4552}{\text{Si IV } 4089}$ $\begin{cases}<1\\=1\\>1\end{cases}$	$\frac{\text{Bl } 4650}{\text{He II } 4686}$ $\begin{cases}\frac{\text{N II } 3995}{\text{He I } 4009}\\ \frac{\text{He I } 4144}{\text{He I } 4121}\\ \frac{\text{O II } 4416}{\text{He I } 4144 \text{ or } \text{He I } 4387}\end{cases}$
B2, B3, B5, B8: $\frac{\text{Si II}}{\text{He I}}$ (Si II 4128, He I 4121, Si II 4128, He I 4144)	H I Balmer lines
B9: $\frac{\text{Ca II K}}{\text{He I}}$ $\frac{\text{He I } 4471}{\text{Mg II } 4481}<1$	$\frac{\text{He I } 4121}{\text{He I } 4144}$, $\frac{\text{N II } 3995}{\text{He I } 4009}$
A0, A1, A2: weak He I lines, Fe present; $\frac{\text{Fe I, Ca I, Mn I}}{\text{Mg II } 4481}$ $\begin{cases}\text{Fe I } 4385\\ \text{Mn I } 4032 \text{ Bl}\end{cases}$	
A3: $\frac{\text{Bl } 4032}{\text{Bl } 4130}$, $\frac{\text{Bl } 4300}{\text{Fe I } 4385}$	$\frac{\text{Bl } 4128-4132}{\text{Bl } 4171\ \ 4179}$
	$\frac{\text{Ca I } 4227}{\text{Mg II } 4481}$, $\frac{\text{Bl } 4416}{\text{Mg II } 4481}$, $\frac{\text{Bl } 4175}{\text{Bl } 4030-4034}$
A5, A7, F0, F2: $\frac{\text{Bl } 4030-4034}{\text{Bl } 4128-4132}$; CH band	$\frac{\text{Bl } 4416}{\text{Mg II } 4481}$, $\frac{\text{Bl } 4416}{\text{CII } 4300}$
F5, F6, F8, G0, G2, G5, G8: $\frac{\text{Fe I or Ca I}}{H_\delta \text{ or } H_\gamma}$; $\frac{\text{Mn I } 4030}{\text{CH } 4300}$	$\frac{\text{Bl } 4172}{\text{Ca I } 4226}$, $\frac{\text{Sr II } 4077}{\text{Fe I } 4045}$, $\frac{\text{Sr II } 4077}{\text{Ca I } 4226}$, $\frac{\text{Bl } 4171-4173}{\text{Ca I } 4226}$, $\frac{\text{Sr II } 4077}{\text{Ca I } 4226;\ H\gamma}$, $\frac{\text{Sr II } 4077}{\text{Ca I } 4226;\ 4063:\ \text{Fe I } 4144,\ 4085}$, $\frac{\text{Ca I } 4226}{\text{Fe I } 4045}$
K0: $\frac{\text{Bl } 4290}{\text{CH } 4300}$	Continuum near 4215 Å (CN bands)
K2, K3, K5: $\frac{\text{Ca I } 4226}{\text{Fe I } 4325}$, $\frac{\text{Bl } 4290}{\text{CH } 4300}$	
M: TiO bands	$\frac{4045}{4077}$, $\frac{4215}{4250}$, $\frac{4376}{4383}$, $\frac{4383}{4390}$

* 1) In the form used by J. Pecker, E. Schatzman. See also
2) MKK classification: W.W. Morgan, P.C. Keenan, and E. Kellman.—Atlas of Stellar spectra, with an outline of spectral classification. Chicago, 1943.
3) MK classification: W.W. Morgan and H. L. Johnson.—ApJ, 117, 313, 1953; Ch. Fehrenbach: Les classifications spectrales des etoiles normales.—Hd. d. Ph. Bd. L, 1—92. 1958.
4) Bl in the table designates a blend (overlapping lines), the wavelengths are in Å.

In addition to the preceding it should be kept in mind that the Harvard HD classification is nowadays used on a progressively smaller scale. This linear classification has been replaced by the two dimensional (Sp, M_v) classification, designated as MKK in honor of its originators Morgan, Keenan, and Kellman, who first published it in their Atlas of Stellar Spectra (1943). It was subsequently improved by a group of Yerkes and other astronomers (Morgan, Johnson, Bidelman, Roman, and others) and was briefly designated as MK for spectral types O5—M2. The luminosity classes Ia, Ib, II, III, IV, V were introduced (see above). This classification of stellar spectra is therefore inseparably linked with the luminosities. Since it was not included in Vol. I, we list here its fundamental criteria for the determination of spectral type and luminosity. They are designed for low-dispersion spectrograph, so as to make stars of all subclasses accessible. The main (approximate) criteria of the MK classification are listed in Table 16. Intensity ratios of lines and blends (Bl) are mainly used. A whole range of objects do not satisfy these criteria, in particular, WR stars, ML stars with enhanced metallic lines, stars with high proper motions, stars from the halo component of the Galaxy, stars with peculiar spectra, etc.

Table 17 lists corrections (from the regression curve) for conversion from MK to HD system. To first approximation, a linear dependence is observed.

TABLE 17*

Sp (MK)	Correction (in tenths of a class)	Sp (MK)	Correction (in tenths of a class)	Sp (MK)	Correction (in tenths of a class)
O6	+0.6	A0	0.0	G0	−0.4
O8	+0.2	A2	−0.4	G2	−0.8
O9	+0.2	A3	−0.7	G5	−0.6
B0	+0.1	A5	−0.2		
B1	0.1	A5	−0.5	K0	0.0
B2	+0.3	F0	−0.3	K3	−1.2
B3	0.0	F2	−0.5	K5	0.0
B4	−0.1	F5	−0.1	K8	0.0
B5	−0.1	F8	+0.4	M0	+0.5
B8	−0.2				

* See analogous Table 13 (approximate).

Photometric methods of integrated-light colorimetry, which are being developed in recent years, will enable the spectral types and luminosities to be determined from narrow bands in the continuous spectra of stars, in particular from photoelectric data. A certain modification of these methods calls for determination of spectral types from color indices, etc.

A two-dimensional spectral classification is being developed for low-dispersion prism cameras — of the order of hundreds Å per mm (Warner and Swasey Observatory and Harvard Observatory in the USA, Crimean and Abastumani observatories in the USSR, etc) and for ultralow dispersion, of the order of tens of thousands Å per mm (Yerkes Observatory in the USA and Pulkovo in the USSR).

In conclusion of this review of absolute magnitude determination techniques note that high-precision determination of spectroscopic parallaxes from from spectral lines is limited only by the feasibility of obtaining slit spectra with dispersions of 10—40 Å/mm. For a 5-m telescope with 1 hr exposure, the corresponding limit is 12 , and for higher exposure times even fainter stars can be processed. One must not forget, however, that this does not detract from the importance of trigonometric parallaxes. After all, any astrophysical method reduces in the final analysis to calibration of spectroscopic parallaxes from the trigonometric parallaxes of the nearest stars.

O.A. MEL'NIKOV

Chapter II

STELLAR MASSES

§8. General. There are known stars with masses ranging from 0.008 to 400 solar masses, but the most frequent ones have their masses between 0.4 and 4.0 of the solar mass. Remember that the mass of Jupiter is $0.00096\,\mathfrak{M}_\odot \approx 0.001\,\mathfrak{M}_\odot$. Our knowledge of stellar masses is derived almost exclusively from observations of binary and multiple systems. Particularly reliable results are obtained for visual binaries. These remarkable objects were originally studied visually, although now they are tackled by photographic and interferometric methods.

Unfortunately, observations of visual binaries provide adequate information only on stars of average and moderate luminosity and average mass, since it is these stars that are mostly observed in binary systems. Moreover, these are mainly late-type stars. Studies of spectroscopic binaries, on the other hand, supply information on objects of high and medium luminosities and masses. Moreover, Algol-type eclipsing variables show secondary components which lie above the main sequence in the Hertzsprung Russell diagram and yet below the giants branch. These are highly remarkable objects called s u b g i a n t s. The primary components of these systems, however, are normal main-sequence stars. Eclipsing binaries also include W UMa stars — close systems of A—K-type stars with anomalous masses.

Detailed formulas for the component masses of binary stars are derived in §11. Here we consider this problem only briefly.

Eq. (95) in Chapter I was

$$\mathfrak{M}_1 + \mathfrak{M}_2 = \frac{a^3}{P^2},$$

where the distance between the components a is expressed in astronomical units. Measurements of binary systems give a in angular seconds (a'') and not in astronomical units. If r is the distance, we have $a/r = a'' \cdot \sin 1''$, i.e., $a'' = \frac{a\,(\text{a.u.})}{r\,(\text{a.u.}) \cdot \sin 1''} = \frac{a\,(\text{a.u.})}{r\,(\text{pc})} = a(\text{a.u.})p$. Hence, if the parallax is known, we obtain

$$\mathfrak{M}_1 + \mathfrak{M}_2 = \frac{(a'')^3}{P^2 (\rho'')^3}. \qquad (1)$$

The value of this sum is generally close to 2. If the relative orbit, i.e., the position of the satellite relative to the major component of the binary

system, is known, we can obtain only the sum of the two masses. If the absolute orbit is known, then $a=a_1+a_2$ and we have

$$a_1 \cdot \mathfrak{M}_1 = a_2 \cdot \mathfrak{M}_2. \tag{2}$$

Moreover,

$$a_1 \frac{\mathfrak{M}_1}{\mathfrak{M}_\odot} = a_2 \frac{\mathfrak{M}_2}{\mathfrak{M}_\odot}. \tag{3}$$

This case corresponds to a known absolute orbit, when observations of the components relative to the background of fixed stars give their individual motions relative to the center of gravity of the system.

For spectroscopic binaries, a is obtained in kilometers. However, a enters all the formulas together with the sine of the inclination angle sin i. The inclination i is known only if the system is at the same time an eclipsing variable binary (then in most cases sin $i \approx 1$).

In general for spectroscopic binaries we have

$$(\mathfrak{M}_1+\mathfrak{M}_2)\sin^3 i = \frac{(a_1 \sin i + a_2 \sin i)^3}{P^2} \tag{4}$$

and moreover

$$(\mathfrak{M}_1+\mathfrak{M}_2)\sin^3 i = \frac{P(1-e^2)^{3/2}}{4G\pi^2}[(\alpha_1+\beta_1)+(\alpha_2+\beta_2)], \tag{5}$$

$$\frac{\mathfrak{M}_1}{\mathfrak{M}_2} = \frac{a_2}{a_1} = \frac{\alpha_2+\beta_2}{a_1+\beta_1}. \tag{6}$$

Here α_1, α_2 and β_1, β_2 are the extreme values of the radial velocities of the two components on the velocity—phase curve.

From equations (5) and (6) we obtain the individual masses, provided that i is known and the velocity curves of the two components of the spectroscopic binary (with split spectral lines) have been measured. In this case α_1 and α_2 are computed from the maximum and β_1, β_2 from the minima of the radial velocity curves of the first and the second component, respectively. For example, for u Her $P = 2^d.051$, $e = 0.05$ (the eccentricity). The star is a spectroscopic and at the same time eclipsing binary (therefore sin $i \approx 1$). For this star $\alpha_1 = 65$, $\alpha_2 = 240$, $\beta_1 = 125$, $\beta_2 = 300$ km/sec and from the above formulas we have $\mathfrak{M}_1+\mathfrak{M}_2 = 11.6$, $\mathfrak{M}_1 = 8.6$ and $\mathfrak{M}_2 = 3.0$. Taking radii 5.28 and 3.67 $R_\odot$ and spectral types B3 and F0, we obtain from the mass—luminosity relation 6.18 and 2.48.

For spectroscopic binaries with an unsplit spectrum we can determine only a function of the component masses $\frac{\mathfrak{M}_2^3 \sin i}{(\mathfrak{M}_1+\mathfrak{M}_2)^2}$ (it has the dimension of mass, if absolute, not relative, masses are used). If i is known, only the lower limit value of the mass can be computed. If, however, a large sample of binaries have been studied and all the orbital inclinations are equiprobable, then their number in an interval di is $\sim 2\pi \sin i \cdot di$ and on the average $\overline{\sin^3 i} = \frac{\int_0^{\pi/2} 2\pi \sin^4 i \cdot di}{\int_0^{\pi/2} 2\pi \sin i \cdot di} = \frac{3}{16}\pi = 0.59$. A somewhat higher value (2/3) is generally assumed, since binaries with large i show a negligible line shift and generally remain undetected (on account of measurement errors).

Spectroscopic and eclipsing binaries are studied spectroscopically and photometrically. Measurements of visual binaries are carried out on an equatorial with a micrometer, on meridian transit instruments, on interferometer, or even photographic negatives. The equations for the measured rectangular coordinates (a geocentric orbit) used in the reduction of measurements are

$$X = c_x + \mu_x t + \pi \cdot P_\alpha,$$
$$Y = c_y + \mu_y t + \pi \cdot P_\delta. \qquad (7)$$

Here c_x and c_y are the heliocentric positions at the zero epoch, e.g., 1950.0, μ_x and μ_y are the annual proper motions, t is the time in years from the zero epoch, π is the relative parallax, P_α is a function of the radius-vector of the Earth's orbit (in A. U.) and right ascension, reduced to a great circle, and P_δ is a function of the radius-vector of the Earth's orbit, right ascension and declination (P_α and P_δ are the parallax factors). The latest data are arranged in Table 1 on the basis of van de Kamp's review (Hd. d. Ph., L, p. 208. 1958) for 12 visual binaries with accurately known orbit elements, mass ratios, parallaxes, absolute magnitudes, and spectra. Highly important information can be obtained by combining spectroscopic and photometric observations with astrometric measurements (Olden, 1946, and others). Suitable cases are αUMi, ηPeg A, 58 Per, and especially Procyon (Strand, 1951). Algol with three components A, B, C has also been studied. The components A and B of Algol constitue an eclipsing binary ($P_{A,B} = 2^d.867$). The period associated with the motion of component C is $P_C = 1.873$ years (determined spectroscopically). The orbit semiaxis (van de Kamp, 1951, and others) $a = 0''.015 = 0.35$ A. U. was measured astrographically relative to the photocenters of AB and C. Given the mass $\mathfrak{M}(A+B) = 6\mathfrak{M}_\odot$, we have $\mathfrak{M}(C) = 1.45\ \mathfrak{M}_\odot$ (for $i = 63°$).

TABLE 1

Star	a''	P (years)	p''	$\mathfrak{M}_1+\mathfrak{M}_2$	$\mathfrak{M}_1$	$\mathfrak{M}_2$	Sp_1	Sp_2	M_b	
									1	2
									m	m
η Cas	11.99	480.0	0.170	1.52	0.94	0.58	G0V	K5+	4.54	+7.51
o^2 Eri B, C	6.894	247.92	0.201	0.66	0.45	0.21	B9	M5e	10.26	9.5
Ross 614A, B	0.98	16.5	0.251	0.22	0.14	0.08	dM6+	—	9.9	11.9
Sirius	7.62	49.94	0.379	3.26	2.28	0.98	A1V	A5	0.80	11.22
Procyon	4.55	40.65	0.287	2.41	1.76	0.65	F5 IV—V		2.59	12.62
α Cen A, B	17.655	80.09	0.760	1.96	1.08	0.88	G4	K1	4.40	5.65
ξ Boo	4.884	149.95	0.148	1.60	0.85	0.75	G8V	K5	5.41	6.70
ζ Her	1.35	34.42	0.104	1.85	1.07	0.78	G0IV	dK0	2.94	5.52
Fu 46	0.71	13.12	0.155	0.56	0.31	0.25	M4	M4	8.26	8.64
70 Oph	4.551	87.85	0.199	1.55	0.90	0.65	K0V	K4	5.56	6.85
Krü 60	2.412	44.6	0.253	4.36	0.272	0.164	dM4	dM6	9.11	9.97
85 Peg	0.83	26.27	0.080	1.62	0.82	0.80	G2V	—	5.26	7.18
Sun	—	—	$0.2063 \cdot 10^6$	—	1	—	G2V	—	4.62	—

Other eclipsing binaries which have been studied are VV Cep and ε Aur (van de Kamp, 1958). Astrometric and astrographic studies of dark (invisible) stellar satellites of small mass are of considerable interest; these include the component of Ross 614 B, the component of L 726 = 8, the component of 61 Cyg with masses of 0.08, 0.04, and 0.008 $\mathfrak{M}_\odot$,

respectively, and others (Strand, Deich, and others). In the case of 61 Cyg we are very close to the mass of Jupiter 0.00096 $\mathfrak{M}_\odot$, so that the mass limits of stars and planets are not clearly separated. There are probably some planets in the Universe whose masses are greater than the currently known maximum and stars with still smaller masses (up to the theoretical limits).

§9. Determination of mass from the red shift. Fairly accurate observation material is currently available for the determination of masses from the red shifts of spectral lines predicted by the general theory of relativity. These measurements make it possible to calculate the masses of isolated, single stars. Escaping from a star of mass $\mathfrak{M}$ and gravitational potential $\Phi = G\frac{\mathfrak{M}}{R}$ into the interstellar space with velocity c, a quantum of light $h\nu_0$ of mass* $m_q = \frac{h\nu_0}{c^2}$ loses energy (does work) equal to its potential energy $h\nu_0 - h\nu = G\frac{m_q\mathfrak{M}}{R} = \Phi\frac{h\nu_0}{c^2}$. The energy $\Phi\frac{h\nu_0}{c^2}$ is lost and dissipated in space, and we therefore observe a quantum of energy $h\nu = h\nu_0 - \Phi\frac{h\nu_0}{c^2}$. Hence $\nu_0 - \nu = \frac{\Phi\cdot\nu}{c^2}$ or equivalently $\frac{\Delta\lambda}{\lambda} = \frac{\Phi}{c^2} = G\frac{\mathfrak{M}}{Rc^2}$. Expressing $\mathfrak{M}$ and R in solar units, we get $\Delta\lambda = \Delta\lambda_\odot\frac{\mathfrak{M}}{R}$ and

$$\frac{\Delta\lambda}{\lambda} = 2.115\cdot10^{-6}\frac{\mathfrak{M}}{R}. \qquad (8)$$

Expressing the red shift in terms of velocity, we obtain

$$\Delta v = 0.635\frac{\mathfrak{M}}{R}\frac{\text{km}}{\text{sec}}. \qquad (9)$$

Since the gravitational potential at the Earth's surface is markedly less than on the surface of the star, the red shift is actually observable. Eq. (9) can be directly used. The results for individual stars are not particularly reliable. The only exceptions are the white dwarfs and the massive O-type stars. Bottlinger (1931) correspondingly used the radial velocity of 286 B-type stars of small peculiar velocities. He subtracted the solar motion and the galactic rotation from these figures and obtained the group values of $\frac{\mathfrak{M}}{R}$. R was calculated from luminosities and temperature, and he obtained for the masses $\mathfrak{M} = 78-1200$. Later Lundmark established that the masses computed in this way are correlated with luminosities. A similar technique can be applied to multiply systems and O-type stars of high mass and luminosity from galactic cluster. Trumpler (1935) carried out the corresponding calculations. His results are listed in Table 2.

The radii and the densities $\varrho/\varrho_\odot$ were estimated from luminosities and temperature. Large line shifts of the order of 20 km/sec should be expected in the spectra of white and other dwarfs (DB, DA, DF, etc., or alternatively wdB, wdA, wdF...; according to another notation system B VIIa, A VIIa, F VIIa (with nuclear sources of energy) and B VIIb, A VIIb, F VIIb (without nuclear energy sources)). Indeed, according to Popper (1954), the white dwarf 40 Eri B of spectral type DA, luminosity class VII ($M_v = +11^m.0$), parallax $p'' = 0''.200$, color $(B-V) = +0^m.03$ and $(U-B) = -0^m.70$ shows a red shift of $(+21\pm4)$ km/sec. This corresponds to $\frac{\mathfrak{M}}{R} = 33.1$. For $R = 0.0132$ (Greenstein, 1948), this gives $\mathfrak{M} = 0.44$.

* The inertial mass of a photon is determined from the relation $e_q = h\nu_0 = m_q \cdot c^2$, i.e., $m_q = \frac{h\nu_0}{c^2}$.

TABLE 2

Cluster	Number of stars measured	Red shift, corrected for reference stars	$\frac{\mathfrak{M}/\mathfrak{M}_\odot}{R/R_\odot}$	Sp	$T\cdot 10^{-4}$ °K	M_{pg}	$R/R_\odot$	$\mathfrak{M}/\mathfrak{M}_\odot$	$\frac{\varrho}{\varrho_\odot}$
NGC 2244	15	+7.9	12.5	O6	4.2	−4.4	6.8	8.5	0.27
	8	8.5	13.4	O9	2.85	−4.1	7.1	90	0.25
NGC 2264	60	14.6	23.0	O7	3.62	−4.7	7.8	180	0.38
NGC 2362	1	9.9	15.6	O8.5	3.02	−6.4	19.5	300	0.04
NGC 6871	2	9.4	14.8	O9w	2.85	−5.7	14.8	220	0.07
	5	15.4	24.0	B0	2.60	−5.4	14.1	340	0.12
NGC 7380	1	+5.4	8.5	B9	2.85	−4.6	8.9	75	0.11

The masses of three white dwarfs which are components in binary systems were determined directly: αCMa B, $\mathfrak{M} = 0.98$, αCMi B, $\mathfrak{M} = 0.4$, 40 Eri B, $\mathfrak{M} = 0.44$. For 40 Eri B, Artyukhina (1948) obtained $\mathfrak{M} = 0.41$ from observations (after orbit correction). These are low-mass stars. Conversely, O-type stars in clusters (Trumpler) have $\mathfrak{M} \approx 300$, i.e., large masses. The reliability of these high stellar masses in clusters, however, has not been fully established. Figure 14 plots the bolometric luminosity as a function of log mass according to the data of Table 16 from Chapter I. The scatter of the points is clearly substantial.

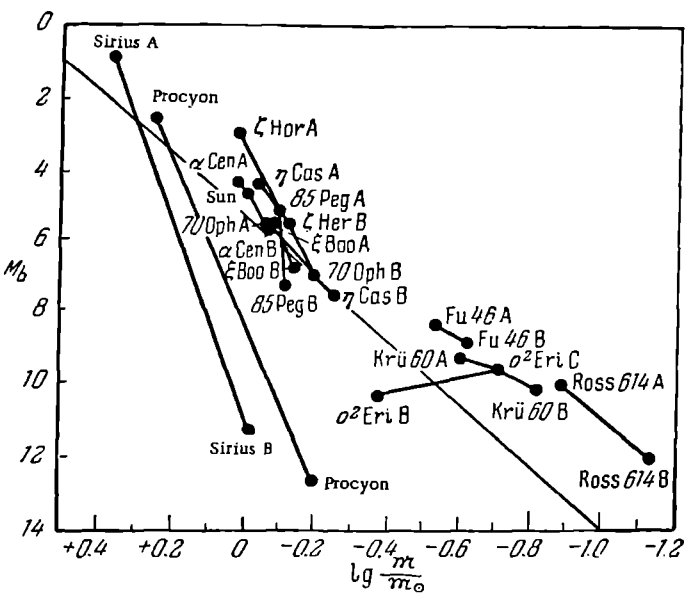


FIGURE 14

In Figure 15 the mass—luminosity relation is plotted from all the data available in 1940, collected in the extensive study and catalogue of H. Russell and Ch. Moore. Data on visual, spectroscopic, and eclipsing binaries and other information were used. We see that the dependence M_b vs. lg $\mathfrak{M}$ is

virtually linear: $M_b = -8.5 \lg \mathfrak{M} + 0.63$ or alternatively $L \sim \mathfrak{M}^{3.4}$, i.e. in practice $L \sim \mathfrak{M}^{10/3}$. After some improvements $L \sim \mathfrak{M}^{3.82}$. Earlier, in 1936, P. P. Parenago obtained $\lg \mathfrak{M} = 0.488 - 0.1009\ M_b + 0.0053\ M_b^2 - 0.00066\ M_b^3 \approx 0.590 - 0.1194\ M_b$, i.e., again approximately $L \sim \mathfrak{M}^{10/3}$. This is the famous mass—luminosity relation. It was first established by Hertzsprung in 1919, and later by Adams, Joy, and also by Eddington. Their results clearly pointed to the great importance of the mass—luminosity relation for the study of the interior stellar structures. In 1940 Russell and Moore obtained $\lg \mathfrak{M} = 0.1048\ (M_b - 5.23)$ or alternatively $\lg L = 3.82 \lg \mathfrak{M} - 0.24$. (Here and in what follows the masses are expressed in units of solar mass.)

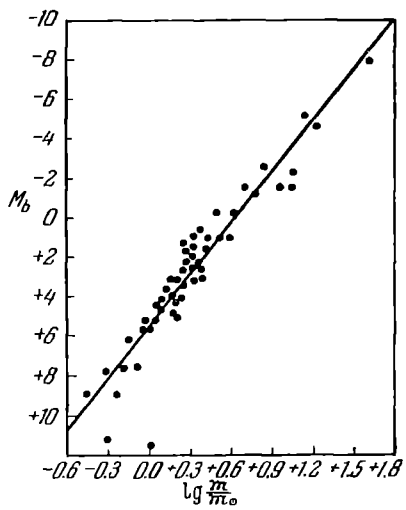


FIGURE 15

We should stress that a definite relation between stellar masses and peculiar radial velocities $\bar{V}_p$ was also discovered, namely $\mathfrak{M} V_p^2 \approx \text{const}$. Unfortunately, in the Galaxy this relation is satisfied only to some approximation. It was derived for the "star gas" by analogy with the motion of gas molecules and atoms.

§10. Mass—luminosity relation. This topic was discussed, in particular, by Masevich and Parenago (1951). They established from observational data that a more general relation is that between mass, luminosity, and radius and that no uniform dependence could be found for stars of all the different types and sequences combined. The mass—luminosity—radius relation was written in the form

$$\lg L = x + y \lg \mathfrak{M} + z \lg R. \qquad (10)$$

The parameters x, y, z were tabulated and found to be different for different stellar sequences, as we see from Table 3. This table also gives the variance σ_M.

TABLE 3

Sequence	Number of equations	x	y	z	σ_M
Main sequence O—G4	19 average values	−0.05±0.08	+1.98±0.60	+2.64±0.80	±0.56
		+0.05 0.09	+3.92 0.17	— —	0.69
		−0.10 0.10	— —	+5.19 0.23	0.71
Main sequence G7—M	13 average values	−0.37 0.06	+1.79 0.17	+1.06 1.06	0.36
		−0.39 0.06	+2.29 0.17	— —	0.37
		−0.37 0.08	— —	+4.49 0.48	0.51
Above the main sequence	5 average values	−0.65 0.08	+1.45 0.29	+4.18 0.60	0.17
		−0.18 0.20	+2.89 0.83	— —	0.71
		−0.76 0.25	— —	+6.34 1.26	0.52
White dwarfs	3 single stars	−4.7	+0.5:	−1.5:	
Subdwarfs	25 single stars	+0.79 0.09	+1.64 0.25	+2.52 0.25	1.12
		+0.71 0.21	+2.13 0.57	— —	2.50
		+0.64 0.15	— —	+2.84 0.42	1.83
Subgiants	26 single stars	−0.35 0.13	+0.33 0.16	+2.05 0.21	0.49
		+0.87 0.09	+0.36 0.37	— —	1.14
		−0.33 0.14	— —	+2.06 0.22	0.52
Giants	9 average values	+0.44 0.07	+0.99 0.54	+0.83 0.13	0.20
		+0.12 0.54	+3.38 1.03	— —	0.52
		+0.78 0.12	— —	+0.99 0.11	0.23
Supergiants	10 single stars	−0.13 0.35	+2.79 0.32	+0.28 0.14	0.75
		−0.15 0.43	+3.18 0.32	— —	0.94
		+2.18 0.76	— —	+1.03 0.37	2.58

Main-sequence stars are clearly divided into two groups with the partial relations

$$1)\qquad L = \underset{\pm 0.23}{1.12}\,\mathfrak{M}^{3.92\pm0.12};\quad \sigma_M = \pm 0.69 \tag{11}$$

for types O—G4;

$$2)\qquad L = \underset{\pm 0.06}{0.41}\,\mathfrak{M}^{2.29\pm0.17};\quad \sigma_M = \pm 0.37 \tag{12}$$

for types G7—M;

$$1')\qquad L = \underset{\pm 0.18}{0.80}\,R^{5.19\pm0.23};\quad \sigma_M = \pm 0.71 \tag{13}$$

for types O—G4;

$$2')\qquad L = \underset{\pm 0.08}{0.43}\,R^{4.49\pm0.48};\quad \sigma_M = \pm 0.51 \tag{14}$$

for types G7—M.

(In (11)—(14), L, R, and $\mathfrak{M}$ are expressed in solar units).

The general relations thus have the form*

$$\begin{aligned} L &= \underset{\pm 0.16}{0.89}\,\mathfrak{M}^{1.98\pm0.60}R^{2.64\pm0.80};\quad \sigma_M = \pm 0.56\ (\mathrm{O}-\mathrm{G4}),\\ L &= \underset{\pm 0.06}{0.43}\,\mathfrak{M}^{1.79\pm0.17}R^{1.06\pm1.06};\quad \sigma_M = \pm 0.36\ (\mathrm{G7}-\mathrm{M}). \end{aligned} \tag{15}$$

The division into two subsequences is confirmed by the character of stellar motions, data for O and T associations (Ambartsumyan, 1947), and other findings.

* Also $\lg R = 0.735 \lg \mathfrak{M} + 0.036$ (for O—G4) and $\lg R = 0.506 \lg \mathfrak{M} - 0.004$ (for G7—M), i.e., $R \sim \mathfrak{M}$.

Lack of a single relation between L, R, and $\mathfrak{M}$ indicates that no single model of stellar structure exists in nature, which could explain the infinite variety of stellar structure and evolution throughout the Universe. Stars are as inexhaustible as the atom. Even now we are constantly encountering various substantial deviations from the above mass—luminosity relations. This was first noted by Mel'nikov (1944) in the study of cepheid masses, which were found to be equal to a few solar masses regardless of their period, i.e., luminosity. Subsequently,this result was confirmed by P. P. Parenago (1955).

Substantial deviations were also observed in eclipsing binaries. These stars were classified by Z. Kopal (1955) and others. Kopal introduced the following three classes (groups):

1) Detached main-sequence systems. The two components statistically follow a relation between L and $\mathfrak{M}^{\alpha}$ ($\alpha = 2.8$ for heavy stars ($\mathfrak{M} > 2$) and $\alpha = 5.7$ for light stars ($\mathfrak{M} \ll 2$)). The partial relation $\mathfrak{M} \sim R^{\beta}$ is also satisfied, with $\beta = 1.5$ for heavy and 1.0 for light stars.

2) Semidetached systems (Algol type). The first component follows the same relations between $\mathfrak{M}$, L and $\mathfrak{M}$, R as the stars of the 1st group. The second component lies above the main sequence and its mass is smaller than that normally corresponding to its luminosity. These stars are known as subgiants. The first, heavier component is normally a smaller and hotter star than the second component. Hence it follows that the principal minima of these systems are the result of total eclipse.

3) Contact systems. These systems occur more frequently than all the others. They belong to spectral types A—K. A typical representative is W UMa. Statistically stars of these systems do not follow the mass—luminosity relation. The first, heavier component of these systems is larger, cooler, and often fainter (in absolute terms) than the second satellite component. The principal minima in these systems are the result of occultation. Note that a more universal classification of eclipsing binaries into 5 groups was advanced by S. Gaposchkin in 1958.

According to O. Struve (1948, 1958), it has been firmly established that the following binaries do not follow the mass—luminosity relation:

a) Weak components of O-type spectroscopic binaries with split lines. Examples, Plaskett's star HD 47129 ($P = 14^{d}.4$) and AO Cas ($P = 3^{d}.52$).

b) Weak components of Algol-type systems with virtually unsplit lines. In some cases they can be observed at the time of the principal minimum.

From the corresponding relations $\left(\alpha = \frac{\mathfrak{M}_1}{\mathfrak{M}_2}\right)$

$$\left.\begin{aligned} \alpha(1+\alpha^2) &= 9.66 \cdot 10^{6} \frac{\mathfrak{M}_1 \sin^3 i}{K_1^3 P (1-e^2)^{3/2}}, \\ f_1(\mathfrak{M}) &= \frac{\mathfrak{M}_2^3 \sin^3 i}{(\mathfrak{M}_1+\mathfrak{M}_2)^2} = \frac{1}{2\pi G} K_1^3 P (1-e^2)^{3/2} \mathfrak{M}_{\odot} \end{aligned}\right\} \qquad (16)$$

it follows that $\mathfrak{M}_2$ is too small or $\mathfrak{M}_1$ is too large (for the given luminosity). Since the main, brighter component has a normal spectrum, the anomaly is probably associated with the second component, a subgiant. For example, for XZ Sgr, Sahade (1949) obtained $f_1(\mathfrak{M}) = 0.004$ and if for the primary normal A3 component we take $\mathfrak{M}_1 = 3$, then $\mathfrak{M}_2 = 0.35$. Hence from the mass—luminosity relation it follows (Kuiper, 1938, and others) that $\Delta M_v = 7^{m}.3$. This is at variance with the number obtained from the photometric light curve $\Delta M_v = 2^{m}.5$. Hence, the luminosity of the subgiant is

100 times greater than what follows from the mass—luminosity relation. In general subgiants have $\mathfrak{M} = 0.2-1.0$ (O. Struve, Huang Su-Shu, 1958), which is too low for their luminosities and effective temperatures. The deviation of subgiants from the $L, \mathfrak{M}$ relation is statistically a function of the mass ratio (O. Struve, 1954), and is also associated with the fact that the secondary subgiant fills one lobe of the inner contact surface (O. Struve, Huang Su-Shu, 1954) of the binary. For W UMa stars, the components are of close spectral types. For 10 systems from K_1 and K_2 we get $\alpha = \frac{\mathfrak{M}_1}{\mathfrak{M}_2} = 2$, although the threshold curves of radial velocities of normal systems with split lines predict $\alpha = 1.25$. The relation between $\mathfrak{M}$ and L for $\alpha = 2.0$ gives $\Delta M_b = \frac{\lg \alpha}{0.1048} = 2^m.9$. On the other hand, from the component spectral and the light curve $L_1 \approx L_2$, i.e., $\Delta M_b \neq 2^m.9$, so that at least one of the components deviates from the mass—luminosity relation. This is apparently due to the fact that this is a very close system, with a common envelope, so that the two stars continuously exchange energy and matter. Struve points to other instances of departure from the standard relation between $\mathfrak{M}$ and L. For example, for HD 698, $\mathfrak{M}_1 \sin^3 i = 113$ and $\mathfrak{M}_2 \sin^3 i = 45$, which corresponds to very high luminosity, which, however, according to the spectrum (late B type) cannot be greater than -1^m. In these and similar cases, however, there is always possible contribution from perturbations in the radial velocity curve of the spectroscopic binaries, which are very difficult to account for and eliminate satisfactorily.

Since the acceleration of gravity at the star's surface is

$$g = G\frac{\mathfrak{M}}{R^2} \tag{17}$$

and the luminosity (assuming a blackbody radiation) is

$$L \sim R^2 T_e^4, \tag{18}$$

we find

$$g \sim \frac{\mathfrak{M}}{L} T_e^4. \tag{19}$$

Since the effect of the absolute magnitude is observed within a given spectral type, i.e., for $T_e \approx \text{const}$, the mass—luminosity relation is in fact equivalent to a relation between g and L. Indeed, in c A2 supergiants, for example (α Cyg), $g \approx 10\,\text{cm/sec}^2$ for luminosities of $M_v \approx -7^m$ (class Ia), whereas in white dwarfs of the same spectral type DA (40 Eri B) $g \approx \approx 10^7\,\text{cm/sec}^2$ and $M_v \approx +11^m$.

For stellar atmospheres in equilibrium at a pressure P (gas + radiation pressure)

$$-\frac{1}{\varrho}\frac{dP}{dh} = G\frac{\mathfrak{M}}{R^2} = G\frac{f(M)}{R^2}. \tag{20}$$

Introducing the optical depth $d\tau = \bar{\varkappa}\varrho\, dh$, we obtain for the pressure gradient

$$\frac{dP}{d\tau} = G\frac{f(M)}{\bar{\varkappa}R^2}.$$

Since $M = -2.5 \lg L + \text{const} = -2.5 \lg (R^2 \cdot T^4) + \text{const}$, we have

$$\lg \frac{dP}{dh} = \lg f(M) + 0.4M + 4 \lg T + \lg \bar{\varkappa} + \text{const}, \qquad (21)$$

and since for one of the star sequences we approximately have $\mathfrak{M} \sim L^{3/10}$, then

$$\lg f(M) = -\frac{3}{25} M + \text{const} = -0.12M + \text{const}.$$

Therefore, the first and the second term in Eq. (21) for the log pressure vary with M in different directions. Since the contribution from the term $0.4\,M$ is greater, we conclude, assuming the transparency to be independent of M, that the log pressure gradient in the atmosphere decreases with increasing luminosity (i.e., decreasing magnitude M).

Detailed study of this problem can be carried out only after the variation of $\mathfrak{M}$ and $\bar{\varkappa}$ with luminosity L (or magnitude M) has been firmly and conclusively established. Numerous astrophysicists, both theoretical and experimental, are currently engaged on this problem.

Quite recently stellar masses began to be determined from photoelectric recordings of stellar continua (see, e.g., Sky and Telescope, 21:322. 1961). The continuum gives the effective T_e and g_e and, since the abundance of the elements is known, comparison with a theoretical model gives R from T_e, L, and then from g and R one computes $\mathfrak{M}$.

Mel'nikov has further (1948) developed a theoretical method for the calculation of masses of long-period cepheids, assuming their masses to be independent of the period. This method was independently confirmed by the results of P. P. Parenago (1955), who used an entirely different, statistical approach (see Chapter VII).

REFERENCES FOR CHAPTERS I and II

1. ADAMS, W.S., A. JOY, M. HUMASON and BRAYTON. The spectroscopic absolute magnitude and parallaxes of 4179 Stars.— ApJ 81, 187. 1935.
2. BARTAYA, R.A. Determination of absolute spectroscopic magnitudes of weak B5—A7 type stars.—Bull. Abastumani Observ., 15:37. 1953.
3. FEHRENBACH, Ch. Les classifications spectrales des étoiles normales.— H.d.Ph., 50;1. 1958.
4. GALKIN, L.S. Two-dimensional classification of very-low-dispersion stellar spectra.— Izvestiya Crimea Astrophys. Observ., 9;82. 1952.
5. GAPOSCHKIN, S. The eclipsing binaries.— H.d.Ap., 50;225. 1958.
6. Henry Draper Catalogue, Harv. Ann. 91—99. 1918—1924; Henry Draper Extension, Harv. Ann. 100, 105. 1937; 112, 1949.
7. KALANDADZE, N.B. Determination of the absolute magnitudes of weak late-type stars (G and K) from spectra taken with an objective prism.— Bull. Abastumani Observ., 10;63. 1949.
8. KALANDADZE, N.B. and R.A. BARTAYA. Determination of spectroscopic parallaxes up to ninth stellar magnitude.— Astron. Circular, Nos. 101—102;7. 1950.
9. Van de KAMP, P. Visual binaries.— H.d. Ap. 50;225. 1958.
10. KEENAN, P.C. Stars with peculiar spectra.— H.d.Ph., 50;93. 1958; MORGAN, W.W.—ApJ 94;501. 1941.
11. KOPYLOV, I.M. Two-dimensional quantitative spectral classification of 238 O5—B7 stars and construction of spectrum—absolute magnitude diagrams.— Izvestiya Crimea Astrophys. Observ., 20;157. 1958.

12. LUNDMARK, K. Luminosities, colours, diameters, densities, masses of stars.— H.d.Ap., Erste Hälfte, Das Stern System, Erster Teil, S.210. 1932.
13. MORGAN, W.W., A.B. MEINEL and H.M. JOHNSON. Spectral classification with exceedingly low dispersion.— ApJ 120:506. 1954.
14. MORGAN, W.W., P.C. KEENAN and E. KELLMAN. An atlas of stellar spectra (MKK).— Chicago. 1943.
15. PARENAGO, P.P. and A.G. MASEVICH. Investigation of the mass—radius—luminosity relation.— Trudy Shternberg Astron. Institute, XX, pp.1—129 (Part I), pp.130—146 (Part II). 1951.
16. RUSSELL, H.N. and Ch.E. MOORE. The masses of the stars.— Chicago, 1940.
17. STRUVE, O.L. and Su-Shu HUANG. Spectroscopic binaries.—H.d. Ph., 50:243. 1958.
18. WALDMEIER, M. Einfürung in die Astrophysik.— Basel. 1948.
19. WOOLLEY, R. and D. STIBBS. The outer layers of a star.— Oxford. 1953.
20. CHALONGE, D. Continuous stellar spectra and the classification of stars.—Astron. Zh., 33:490. 1956; Classification of stars.—Priroda, No.11:42. 1954.
21. SHAIN, G.A. Absolute magnitudes and masses of stars.— In: "Kurs Astrofiziki i Zvezdnoi Astronomii," Vol.II, p.231. 1936.

Part Two

BINARY STARS

Binary and multiple systems constitute a very extensive class of objects in our Galaxy. As a matter of fact, solitary stars appear to be an exception rather than a rule. Our Sun with its giant massive planets, Jupiter and Saturn, may be regarded as a multiple system, not unlike the recently discovered systems with dark companions whose masses are 10^{-3}—10^{-4} of the mass of the central body.

Binary stars are conventionally divided into three groups. This division, however, is largely based on observation methods, and not always on physical characteristics. The first of these groups includes the visual binaries, discovered by straightforward astrometric observations. Visual techniques are currently supplemented on a large scale by photographic measurements and very occasionally by interference methods. A better name for this group is thus astrometric binaries, and not visual binaries, to emphasize that these binaries are identified by astrometric, rather than photometric or spectroscopic, techniques.

Photometric binaries, also known as eclipsing binaries are investigated by photometric methods. Spectroscopic binaries are characterized by periodic variation of their radial velocities, as observed with a spectrograph. These two methods detect relatively close pairs of stars whose orbital planes do not make a large angle with the line of sight. The astrometric method is suitable for measuring wider pairs, irrespective of the orbital inclination.

A.N. DEICH

Chapter III

VISUAL BINARIES

§11. A historical survey. The first telescopes showed that some of the stars which appeared solitary to the naked eye were unexpectedly resolved into two separate stars. In 1650 Riccioli established the binary character of ζ UMa — Mizar — with separation of 14" between the components. In the years that followed the binary character of Castor, 61 Cyg, α Cen, and other stars was discovered more or less by chance. A few dozen of such stars had been found by 1775. Christian Mayer published in 1781 a list of 89 binary stars from his observations with a mural quadrant in Mannheim; he further claimed to have noticed the motion of the satellite about the primary. This bold assertion was the laughing stock of his contemporaries, and the "planets" revolving around primary stars were regarded as a product of hallucination. Ch. Mayer's measurements were naturally of low accuracy and went on only for two years, so that they could hardly provide conclusive proof of the orbital motion of one star about the other. The idea as such, however, was definitely correct and it fitted a similar suggestion advanced by Goodricke concerning Algol, whose variable brightness Goodricke attributed to periodic occultations by a dark companion revolving around the luminous star. A decade earlier James Michell had proved from probability considerations that the existence of a large number of close binaries could not be a chance factor. In most cases the components of binary systems must be linked by some physical bond. It was William Herschel, however, who first obtained undisputable observational proof of the existence of double stars.

Herschel built his own large reflectors and with great enthusiasm devoted himself to celestial observations. In 1782 he published the first catalogue of 269 double stars. A year later he published his second catalogue listing 703 binaries and in 1821, only a year before his death, Herschel published a third catalogue with 145 new binaries. Herschel measured the separations and the position angles and estimated the magnitude and the color of the components. The catalogues included stars with separation not exceeding 2'. The measurements involved considerable difficulties, as Herschel's reflectors did not have any clock mechanism.

Herschel tried to measure the parallaxes of binaries by putting to work Galileo's idea for the determination of the parallaxes of bright stars relative to weak "optical companions," which are actually much farther than the bright star and are observed in the same direction only by chance. Repeated

measurements, however, led to entirely different results. Herschel thus unquestionably observed the motion of the satellite about the primary in some 50 binaries and derived the orbital periods for some of them. For Castor the period P was found to be 342 years, which is close to the present-day value of 360 years. The dispute around the existence of physical binaries was thus finally settled. The year 1803, when Herschel published his memorandum on the variation of the relative positions of components in binary systems and the possible reasons for this effect, by right should be considered the year of discovery of double stars. The effects of Newton's law of gravitation were for the first time vividly observed outside the solar system, which conclusively established its universal applicability.

William Herschel's son — John Herschel — took up his father's work and together with James South added 380 binaries in the North Hemisphere and 2100 binaries in the South Hemisphere. The list of these new binaries was published in 1847. Like W. Herschel's binaries, these lists included a fair proportion of optical binaries, but most of them were nevertheless physical binaries with very large orbital periods.

Although W. Herschel is rightly considered the discoverer of binary stars, it was W. Struve who emerged as the pioneer of exact and systematic measurements and developed a routine technique for the observations of binaries. Before the establishment of the observatory at Pulkovo, Struve carried out observations of binary stars at Dorpat (now Tartu) in Estonia. The results of the first measurements made with a transit instrument and Trouton's small equatorial were published in 1822 in the form of a catalogue comprising 795 binaries. Encouraged by his success, Struve took decisive steps toward the acquisition of new and better instruments. After his trip to the West and, in particular, his visit to the Koenigsberg Observatory, then directed by the notorious Bessel, Struve purchased a new meridian transit circle and a Fraunhofer refractor with a focal distance of 4 m and aperture diameter of 25 cm, equipped with a clock mechanism. This instrument delivered at Dorpat in 1824 was the largest refractor of that time. Struve was reported to have said that his refractor would compete with any of Herschel's famous reflectors. It was on this instrument that Struve carried out his fundamental research comprising three distinct stages: 1) search for binaries, 2) determination of their equatorial coordinates, and 3) micrometric measurement of the separation and position angles. The magnitudes and colors were also determined. The results were published in three large folios, Catalogus Novus in 1827, Positiones Mediae in 1852 (epoch 1830.0), and Mensurae Micrometricae in 1837. Some 120,000 stars brighter than 9 mag. were examined in the part of the sky from the North Pole to declination −15°. These included 2640 binaries, listed in Mensurae Micrometricae, where they are divided into four classes according to the separation between the components: less than 4", from 4" to 8", from 8" to 16", and from 16" to 32". W. Struve's catalogue has retained its scientific value to this day. The double stars from this catalogue are designated in the literature by Σ. In Chapter XIV of his treatise Struve considers the question of parallax determination and gives the results of the first determination of the parallax of Vega ($p = 0''.125$), which is in remarkably good fit with the present-day value, $p = 0''.121$.

In 1839, a new refractor, the largest in the world for some time, was erected at the Pulkovo Observatory ($f = 7$ m, $d = 38$ cm). This in a sense became the personal instrument of Otto Struve, W. Struve's son, who worked with it for 46 years, an unusually long period for one observer. The program of O. Struve's observations included the previously discovered binaries with considerable relative motions and also close binaries not detected by previous observers, who had used weaker instruments. O. Struve discovered 547 such stars. In addition to direct observations. O. Struve made an extensive study of the biased errors in the measurements of binaries, in particular the observers' subjective errors. The works of O. Struve were published in volumes IX and X of the Pulkovo Annals for 1878 and 1893. The binaries discovered by O. Struve are designated OΣ.

The work of the two Herschels and the two Struves completes the first period in the observations of binaries. Among other observers who worked in the same period we should mention E. Dembowski, who vigorously carried out observations of the Struve stars between 1851 and 1878 in Naples. Dembowski's work was published posthumously in 1883—1884 under the editorship of O. Struve and Schiaparelli.

In the last quarter of the 19th century the prevailing opinion was that all the double stars accessible to contemporary instruments, at least those in the North Hemisphere, had been exhausted. In 1873, however, Burnham in America published a list of 81 binaries found with a 15-cm refractor. A year later, he found another 100 stars. Without slacking his tempo, Burnham proceeded to discover some 1300 new binaries with various instruments, including the largest refractors with 90 cm and 100 cm apertures. His success was due to the fact that all the new binaries were exceedingly close pairs or had very faint companions, and thus escaped earlier detection.

Burnham's pupil, R. Aitken, with Hussey found 4400 new binaries with separations less than 5" in the part of the sky from the North Pole to declination −22° working in 1899 with the 90-cm Lick refractor. Table 1 lists the number of binaries discovered by various observers and the percentage of pairs with separation less than 2". The percentage of close binaries definitely increases in the work of later observers. Of all the stars in W. Struve's catalogue, only 17% showed noticeable orbital motion; this fact was established by Jackson /1/ who analyzed Struve's observations in 1922, i.e., 100 years after his death. If we further omit the 7% which are probably mere optical binaries (to judge from their proper motions), the remaining 76% prove to be physical binaries with orbital periods of thousands of years. O. Struve's catalogue, according to Jackson, contains 32% of binaries with noticeable orbital motion and only some 5% are optical binaries. A concerted observational effort is thus required in order to determine the orbits of the Struve stars. Ludwig Struve, the son of Otto Struve, began his observations in 1885 in Russia on the 76-cm Pulkovo refractor, then the largest in the world. The results were published in the form of a catalogue comprising 750 binaries. In 1892, however, Struve's program had to be greatly curtailed since the large refractor was placed at A. A. Belopol'skii's disposal for the purpose of spectroscopic observations. Small series of double star observations were carried by S. P. Glazenap at the St. Petersburg University Observatory and by P. K. Shternberg and S. N. Blazhko at the Moscow Observatory.

TABLE 1

Observer	Total number of binaries	Percentage of binaries with $\varrho<2''$
W. Herschel	812	4
W. Struve	2640	15
O. Struve	547	40
Burnham	1290	55
Aitken	4432	74
Rossiter	7368	74

As new observational material became available, a need was felt in a comprehensive catalogue of binary stars. Such a catalogue was compiled in 1906 by Burnham. This catalogue contains some 14,000 binaries from the North Pole to $-31°$ declination. The rapid growth in the volume of observational data led Aitken to publish a new general catalogue in 1932. Aitken was more careful in his selection of binaries according to separation and apparent magnitude, following his formula $\lg \varrho \leqslant 2.8-0.2m$, where ϱ is the separation in seconds of arc, m is the total apparent magnitude of the pair. According to this formula, the double stars with $m = 9.0$ listed in Aitken's catalogue should have a separation of $\leqslant 10''$. For stars of 11.0 mag., the separation is $<4''$. Following his criterion, Aitken was forced to drop about one third of the stars appearing in Burnham's catalogue. This new ADS catalogue nevertheless lists 17,180 stars in the part of the sky from the North Pole to polar distance of 120°. The number of binaries discovered in the Southern Sky was less than in the North Hemisphere, as the observations were much fewer. In this century, the great work of Innes, van den Bos, Rossiter, and other astronomers in South African observatories led to the discovery of several thousands of new binaries. The number of known binaries in the entire sky is now approaching 40,000. A new comprehensive catalogue of double and multiple stars is required. Moreover, Aitken's criterion limits the number of binaries, since weak stars do not satisfy this criterion. On the other hand, note that only 10% of all the recorded binaries show definite relative motion and more or less reliable orbital elements could be computed only for 1% of the stars (some 400).

The present-day phase of double star research is characterized by the search for exceedingly weak low-mass companions, and this requires high-precision observations. Of considerable interest are the wide binaries, identified from their common proper motion. Both these problems are successfully tackled by the methods of photographic astrometry. Considerable attention is devoted to star systems of the Trapezium type. As new observational data become available, the significance of statistical techniques increases. The various aspects of the problem are considered in the following sections of this chapter.

§12. **Observations of visual binaries.** The common technique for visual measurements of binaries has not changed much since the time of W. Struve. We are still using a position micrometer with independently moving hair-lines, which enables us to obtain a simultaneous fix of both the primary and its satellite. In visual measurements we measure the position angle θ of

the satellite relative to the primary star and the mutual separation ϱ (Figure 16). Photographic measurements give the rectangular coordinates x and y. Clearly

$$\left.\begin{aligned} x &= \varrho \sin\theta = \Delta\alpha \cos\delta, \\ y &= \varrho \cos\theta = \Delta\delta. \end{aligned}\right\} \qquad (1)$$

Fairly recently, P. Muller at the Strasbourg Observatory /2/ devised a new micrometer where a birefringent crystal (a Wollastone prism) gives a double image of the star. By moving the prism so as to obtain symmetric configurations of star images, one measures the separations and the position angles. In Muller's opinion, the new method is several times more accurate than the hairline technique.

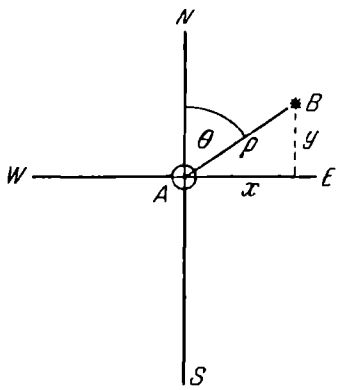


FIGURE 16

The effect of biased errors must be carefully considered in both visual and photographic observations. These errors are determined by a variety of factors, e.g., a difference in the stellar magnitudes of the companions, their mutual separations, the color and spectral type of the stars, etc. Observer's subjective errors are also biased, especially in visual measurements. When observation series carried out by different observers are compared, the subjective errors are eliminated by making a special preliminary comparison between optical pairs chosen as standards or by simultaneously measuring the same physical binaries. The instruments used in visual observations and in photographic measurements introduce a certain error, which may reach hundredths and even tenths of a second of arc. To ensure high-precision determination of orbits, these errors should be detected and eliminated.

Photographic measurements are generally more accurate than the visual measurements, and the biased errors are correspondingly smaller. The visual observations, however, have a definite advantage in measurements of close binaries, since the resolving power of photographic emulsions is markedly less than the resolving power of the human eye; moreover, various photographic effects distort the true separation of the component images. The interference method doubles the resolving power of the visual refractometer, and periscopic attachments even lead to a further improvement. The loss of light, however, is considerable, so that the application of these techniques is limited to stars of up to 5—6 mag., and reliable measurements require exceptionally good atmospheric conditions. A more detailed discussion of the technical aspects of observations of binaries is unfortunately beyond the scope of the present Course.

In conclusion note that when observations of one epoch are converted to another epoch, we must allow for the effect of precession on the position angle θ, and hence on the coordinates x and y, whereas the separation ϱ is not affected. Let P and P' (Figure 17) be the two positions of the pole on the celestial sphere, which has moved in Δt years by a distance $n\Delta t$ toward

the point of vernal equinox γ, where $n = 20''.0 \approx 0°.0056$. In the triangle $PP'A$, the angle at the primary star A is equal to the difference in the position angles $\theta' - \theta = \Delta\theta$ that the line AB (where B is the satellite) makes with the meridians AP and AP'. The angle at the pole P is equal to the right ascension of the star, and the arc AP' is $90° - \delta'$. Clearly,

$$\sin \Delta\theta \cos \delta' = \sin \alpha \sin n\Delta t.$$

The angles being small, we write with fair accuracy

$$\Delta\theta = 0°.0056\, \Delta t \sin \alpha \sec \delta' \qquad (2)$$

The common proper motion of the pair in right ascension also affects the position angle. In Figure 18, the binary AB has moved to the position $A'B'$ in time Δt. The position angles are θ and θ'. Without changing the angle θ', we can move the pair $A'B'$ along the meridian to the point $A''B''$ so that AB and $A''B''$ lie on the same great circle. In the triangle PAA'' the angle at the pole is equal to the proper motion of the star in right ascension multiplied by the time interval, i.e., $\mu_\alpha \Delta t$. The angle $PA''A = 180° - \theta$ and the arc $PA = 90° - \delta$. Then $\cos \theta' = \cos\theta \cos \mu_\alpha \Delta t - \sin\theta \sin \mu_\alpha \Delta t \sin \delta$. The angles being small, we write

$$\cos\theta - \cos\theta' = \sin \mu_\alpha \Delta t \sin\theta \sin\delta$$

or

$$2 \sin \frac{\theta + \theta'}{2} \sin \frac{\theta' - \theta}{2} = \sin \mu_\alpha \Delta t \sin\theta \sin\delta,$$

hence

$$\theta' - \theta = \Delta\theta = \mu_\alpha \Delta t \sin \delta. \qquad (3)$$

In general, this correction is much less than correction (2). For 61 Cyg, a star with one of the largest known proper motions, $\mu_\alpha = 4''.1$ and $\sin \delta = 0.62$. In 100 years, the correction builds up to $0°.07$.

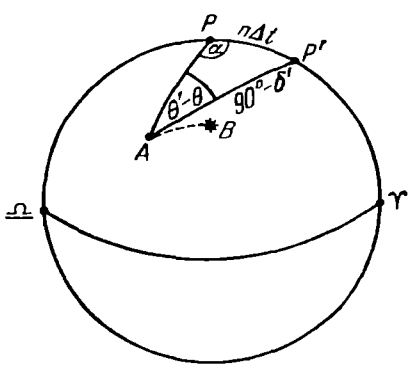


FIGURE 17

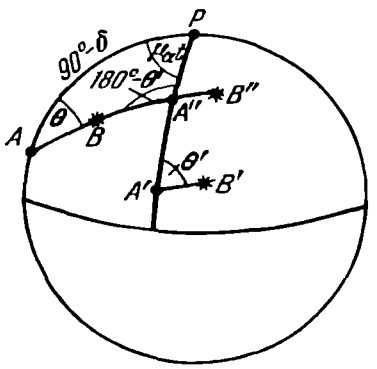


FIGURE 18

§13. Determination of orbits of visual binaries. General considerations. The construction of the apparent ellipse. The true orbit of a binary is always observed in projection onto the tangent plane, perpendicular to the line of sight. On this tangent plane, also called the plane of the sky, we observed the apparent orbit, or the apparent ellipse, as distinct from the true ellipse.

Both stars revolve around their common center of mass, describing coplanar similar ellipses. Mostly we measure the position of the satellite relative to the primary (in photographic astrometry, the motion of each component can in fact be related to reference stars), so that the results give the relative orbit. Evidently, the semimajor axis of the relative ellipse with the primary at its focus is equal to the sum of the semimajor axes of the "absolute" ellipses with their foci at the center of mass of the two stars. The position of the primary, which is situated at the focus of the true ellipse, in general does not coincide with the focus of the apparent ellipse and in fact it may lie anywhere inside the apparent ellipse depending on the projection angle.

The problem of orbit determination of binaries is divided into two parts. First one has to find or construct the apparent ellipse. To this end all observations are reduced to the same epoch, corrections for biased errors are introduced, and the various data are grouped into so-called normal points. The corresponding locus is described by a general equation of an ellipse

$$Ax^2 + 2Hxy + By^2 + 2Gx + 2Fy + 1 = 0$$

and the coefficients can be found by the least squares method. This approach, however, is much too formalistic. The actual times of observation are not taken into consideration here, whereas time is determined with much higher precision than the coordinates and it would be a better policy to use the law of areas, which states that the radius-vector of both the true and the apparent ellipse sweeps out equal areas in equal times. The apparent ellipse is thus constructed graphically, and the construction is checked against the law of areas, which should be satisfied along the entire arc covered by observations. Graphical construction also permits omitting those observations which markedly deviate from the general trend, being distorted by random errors. Note, however, that maximum care should be exercised in omitting the "unpleasant" points, especially if the total number of observations is not large, as in this case the choice of the best ellipse may become quite arbitrary. The observer's objectivity is always an unknown factor and the different weights assigned to individual observations, though related to the accuracy of the measurements, are never quite certain. The construction of the apparent ellipse is a highly responsible and difficult task, which actually determines the accuracy of the elements of the true ellipse. The observations generally cover but a small arc of the orbit, since the orbital periods of the binaries are commonly large. It is not at all easy to draw the complete ellipse using a small arc.

Güntzel-Lingner's technique can be recommended for the construction of the complete apparent ellipse from a small arc /3/. Draw two tangents at the points B_1 and B_2 of the arc, join their intersection point S_1 with the midpoint M_1 of the arc segment between B_1 and B_2. The line S_1M_1 evidently

passes through the center of the ellipse. Two such lines S_1M_1 and S_2M_2 thus intersect at the center of the ellipse C (Figure 19). Points M_1', M_2'..., symmetric to the corresponding points of the observed arc are then laid off on the other side of the center, and the ellipse is drawn through them.

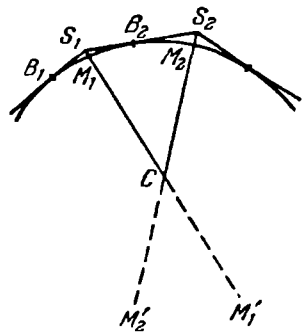


FIGURE 19

Orbital elements. Once the apparent ellipse has been constructed, we can proceed with a determination of the elements of the true orbit. There are seven elements all in all. Two of these characterize the position of the plane or the true orbit, i.e., its inclination i to the tangent plane and the position angle Ω of the intersection line of the two planes. The inclination i is reckoned from 0° to ±90°. The double sign implies that the plane of the true orbit cannot be determined unambiguously, since it is impossible to distinguish the part in front of the tangent plane from the part behind the tangent plane. The radial velocity of orbital motion enables us to determine the sign of the inclination. The position angle Ω of the intersection line, known as the line of nodes, goes from 0° to 180° counterclockwise. * The angle Ω is often called the node of the orbit: nodes in this sense are the intersection points of the true ellipse with the tangent plane, assuming the line of nodes to pass through the primary. If the radial velocity at the node is positive, i.e., the satellite moves away from the observer, the inclination is taken positive.

The orientation of the true ellipse in the orbital plane is characterized by the angle ω between the line of nodes and the major axis of the ellipse, or equivalently by the arc from the node to the periastron, which is measured from 0° to 360° in the direction of satellite's motion.

The other four elements — the so-called dynamic elements — are the orbital period in years P, the time of periastron passage T, the eccentricity of the orbit e, and the semimajor axis a of the satellite's orbit in angular seconds.

Methods of orbit determination. The classical methods of orbit determination of binaries, proposed at the end of the last century, and the more recent techniques developed in this century can be divided into graphical and analytical. Zwiers's graphical method (1895) and the analytical method of Koval'skii (1873), as improved by Glazenap (1889), are among the commonly used techniques, which are described in various textbooks /4, 5/. We will discuss here the graphical method of Mlodzeevskii (1890), which is unjustifiably ignored in the literature, and the Thiele—Innes analytical method (1883—1926), which is fairly often used in current work.

MLODZEEVSKII'S METHOD. Let the apparent ellipse be given, with its foci at F_1 and F_2 and the center O (Figure 20). Draw a tangent MT at the point M, and erect a normal MN; the normal, as we know, will bisect the angle F_1MF_2.

* When viewed from inside the celestial sphere.

The tangent itself bisects the exterior angle F_2ME. The bisectors of the interior and exterior angle divide any line crossing the sides of the angle into proportional parts. We thus get

$$\frac{F_1T}{F_2T}=\frac{F_1N}{F_2N}.$$

We can rewrite this proportion so that it includes the center O of the ellipse. Since

$$\frac{F_1T}{F_2T}=\frac{F_1O+OT}{OT-OF_2}=\frac{OT+F_1O}{OT-F_1O}$$

and

$$\frac{F_1N}{F_2N}=\frac{F_1O+ON}{F_2O-ON}=\frac{F_1O+ON}{F_1O-ON},$$

we have

$$\frac{OT+F_1O}{OT-F_1O}=\frac{F_1O+ON}{F_1O-ON},$$

whence

$$\frac{OT}{F_1O}=\frac{F_1O}{ON}$$

or finally

$$ON\cdot OT=OF_1^2. \qquad (4)$$

We can now proceed with the determination of the elements of the true ellipse. Let as before the apparent ellipse be given with its center O and foci F_1, F_2 (Figure 21). The primary star A, which lies in one of the foci of the true ellipse, and the other focus A' are projected inside the apparent ellipse symmetrically about the center O. Drawing through the point M a tangent MT and a normal MN and joining the points A and A' to the point M, it is no longer true that the normal and the tangent bisect the interior and the exterior angle of the triangle AMA'. In other words, the angle AMS in general is not equal to the angle $A'MT$. The two angles are equal only when the line of nodes is parallel or perpendicular to the tangent ST. This follows from the fact that the equality of these angles in the true ellipse is conserved in projection onto the tangent plane in these two special cases only.

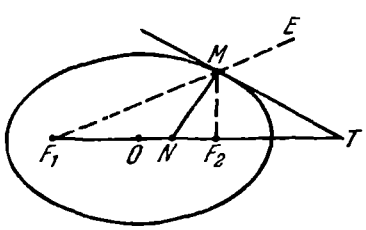


FIGURE 20

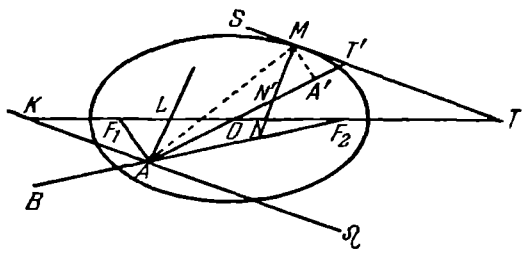


FIGURE 21

Suppose we have found a tangent ST such that the angles at the point M are equal. The intersections of the line AA' with the normal and the tangent are marked N' and T', respectively. Then, from equality (4),

$$ON'\cdot OT'=OA^2. \qquad (*)$$

For the apparent ellipse an analogous equality is always true:

$$ON \cdot OT = OF_1^2. \qquad (**)$$

Through the projection of the primary star A we now draw lines AK and AL parallel to the tangent and the normal, respectively. They meet the line F_1F_2 at points K and L. OAL and ONN' are similar triangles, and thus

$$\frac{OL}{OA} = \frac{ON}{ON'}.$$

From the similar triangles OAK and $OT'T$ we get

$$\frac{OK}{OA} = \frac{OT}{OT'}.$$

Multiplying the last two equalities, we get

$$\frac{OL \cdot OK}{OA^2} = \frac{ON \cdot OT}{ON' \cdot OT}.$$

Using (*) and (**), we finally obtain

$$OL \cdot OK = OF_1^2.$$

This equality shows that the lines AL and AK bisect the interior angle F_1AF_2 and the exterior angle F_1AB, respectively. One of these lines by definition should coincide with the direction of the line of nodes. Thus, to find the line of nodes, it suffices to join the primary star A by straight lines to the foci F_1 and F_2 of the apparent ellipse, construct the angle F_1AF_2, and draw its interior or exterior bisector.

Which of these bisectors is in fact the line of nodes? Consider the case when the primary is projected onto the major axis of the apparent ellipse, outside the segment F_1F_2. The interior bisector then coincides with the major axis and the exterior bisector is perpendicular to it. Since by assumption the foci of the apparent ellipse are closer to the center than the projections of the true foci, the line of nodes is the exterior bisector. Similarly, if the primary lies on the segment F_1F_2, the exterior bisector is the line of nodes and it also coincides with the major axis of the ellipse, since by tilting the ellipse about this axis we increase the eccentricity of the apparent ellipse. If the primary is moved away from the major axis, the exterior bisector will remain the line of nodes, and only its position will change. Thus the line of nodes is always the bisector of the exterior angle forward in the above construction, and the position angle of the line of nodes, or simply the node Ω of the orbit, can be found if the direction of the circle of declination is known.

To find the other elements, consider Figure 22. Here $\Omega A\Omega$ is the line of nodes. The projection of the major axis of the true ellipse is the diameter ab of the apparent ellipse through A, A' and the center O. The projection of the minor axis cd is the conjugate diameter and can be constructed as the line bisecting all the chords parallel to the diameter ab. These two projections, in general, are not perpendicular to each other. Let the diameter cd meet the line of nodes at the point E. We construct a semicircle using AE as the diameter. Dropping the perpendicular OD from the center O

to the line of nodes, we continue it to intersection with the semicircle at the point O'. This is the center of the true ellipse, since the angle $AO'E$, subtended by the diameter AE, is a right angle, and AO' and $O'E$ are directed along the major and the minor axes of the true ellipse. We as if rectified our drawing, having turned it through the inclination angle i around the line of nodes. Clearly, $OD : O'D = \cos i$, whence we find the element i. Dropping a perpendicular from the point b to the line of nodes and continuing it to intersection with AO', we find the apastron b' of the orbit and also the semimajor axis $a = O'b'$. The longitude of the periastron ω is equal to the angle $\Omega A\Pi$. The eccentricity e is obtained as the ratio $O'A : O'\Pi$ or equivalently $OA : Oa$.

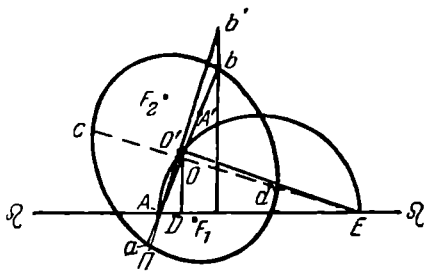


FIGURE 22

The elements P and T also can be found geometrically. Using the law of areas, we calculate the areal velocity, dividing the area of the sector of the apparent ellipse swept out by the projection ϱ of the radius vector by the corresponding time Δt. The area of the sector between individual observations is measured using a drawing of the ellipse on millimeter paper or with a planimeter. Dividing the entire area of the apparent ellipse by the areal velocity, we find the orbital period P of the satellite. The time of periastron passage T is found from observation times by adding or subtracting the times obtained when the area of the sector primary — satellite — periastron is divided by the areal velocity.

THE THIELE—INNES METHOD. Let the apparent ellipse be given. Draw a line through its center and the primary to intersection with the ellipse; this gives the projection of the periastron and the ratio $OA : Oa$ gives the eccentricity e. The elements P and T are found geometrically, as above. To find the other elements, consider Figure 23. The center of the auxiliary sphere is at the primary A. The satellite B moves along the arc of the true orbit, whose plane intersects the sphere along the arc ΩK; $A\Omega$ is the line of nodes.

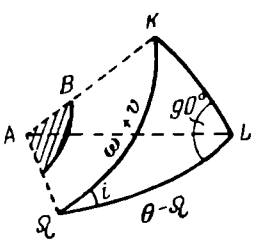


FIGURE 23

The tangent plane makes an angle i with the orbital plane and meets the sphere along the arc ΩL, which is equal to $\theta - \Omega$, θ being the measured position angle of the satellite. The position of the satellite in orbit is characterized by the radius-vector r and the true anomaly v, reckoned from the periastron. The arc ΩK is therefore equal to $\omega + v$. From the spherical triangle ΩKL, where L is a right angle, we have

$$\operatorname{tg}(\theta - \Omega) = \operatorname{tg}(\omega + v)\cos i,$$
$$\cos(\omega + v) = \cos(\theta - \Omega)\cos KL.$$

But

$$\varrho = r\cos KL,$$

so that

$$\varrho \cos(\theta - \Omega) = r \cos(\omega + v)$$

and

$$\varrho \sin(\theta - \Omega) = r \sin(\omega + v) \cos i.$$

Expanding the sine and the cosine of sums and differences of angles, we multiply the first equation by $\sin \Omega$, the second by $\cos \Omega$, and add them up. Then we multiply the first equation by $\cos \Omega$, the second by $\sin \Omega$, and subtract the second equation from the first. This gives

$$\left.\begin{aligned} x = \varrho \sin\theta &= r\cos v(\cos\omega \sin\Omega + \sin\omega\cos\Omega\cos i) - \\ &\quad - r\sin v(\sin\omega\sin\Omega - \cos\omega\cos\Omega\cos i), \\ y = \varrho\cos\theta &= r\cos v(\cos\omega\cos\Omega - \sin\omega\sin\Omega\cos i) - \\ &\quad - r\sin v(\sin\omega\cos\Omega + \cos\omega\sin\Omega\cos i). \end{aligned}\right\} \tag{5}$$

We have

$$r\cos v = a(\cos E - e),$$
$$r\sin v = a\sqrt{1-e^2}\sin E,$$

where a is the semimajor axis of the orbit, e is the eccentricity, and E is the eccentric anomaly. Dividing through by a and introducing a new notation X and Y, we get

$$\frac{r}{a}\cos v = \cos E - e = X,$$
$$\frac{r}{a}\sin v = \sqrt{1-e^2}\sin E = Y.$$

X and Y can be calculated for each observation, since by Kepler's equation

$$E - e\sin E = M,$$
$$M = \frac{360^\circ}{P}(t - T),$$

where t is the time of observation.

Dividing and multiplying the right-hand sides of Eqs. (5) by a, we write

$$\left.\begin{aligned} x &= AX + FY, \\ y &= BX + GY, \end{aligned}\right\} \tag{6}$$

where

$$\left.\begin{aligned} A &= a(\cos\omega\sin\Omega + \sin\omega\cos\Omega\cos i), \\ B &= a(\cos\omega\cos\Omega - \sin\omega\sin\Omega\cos i). \\ F &= -a(\sin\omega\sin\Omega - \cos\omega\cos\Omega\cos i), \\ G &= -a(\sin\omega\cos\Omega + \cos\omega\sin\Omega\cos i). \end{aligned}\right\} \tag{7}$$

To find the unknowns A, F, B, G, which contain the sought elements Ω, ω, i, a, we require only two normal points chosen in the optimal fashion from observations. The problem is also solved by the least squares method using all the observations. The coefficients X and Y are conveniently calculated from the tables published in Union Obs. Circular No. 71, Appendix, 1926.

The sought orbital elements are then calculated as sums and differences of the Thiele—Innes constants, specifically:

$$A + G = 2a\sin^2\frac{i}{2}\sin(\Omega - \omega),$$

$$A - G = 2a \cos^2 \frac{i}{2} \sin (\Omega + \omega),$$

$$B - F = 2a \sin^2 \frac{i}{2} \cos (\Omega - \omega),$$

$$B + F = 2a \cos^2 \frac{i}{2} \cos (\Omega + \omega).$$

Hence

$$\operatorname{tg} (\Omega + \omega) = \frac{A - G}{B + F},$$

$$\operatorname{tg} (\Omega - \omega) = \frac{A + G}{B - F}$$

and we have Ω and ω. Further,

$$\operatorname{tg}^2 \frac{i}{2} = \frac{A + G}{A - G} \frac{\sin (\Omega + \omega)}{\sin (\Omega - \omega)} = \frac{B - F}{B + F} \frac{\cos (\Omega + \omega)}{\cos (\Omega - \omega)},$$

and, finally, the semimajor axis a in angular seconds can be obtained from any of the four equations above, e.g.,

$$a = \frac{A + G}{2} \operatorname{cosec} (\Omega - \omega) \operatorname{cosec}^2 \frac{i}{2}.$$

The elements e, P and T can be improved by comparing the calculated points of the orbit with the observed points. By varying the initial elements e, P, and T, we can minimize the sum of the squares of the deviations from the calculated orbit and adopt the results as the final orbital elements. The position of the satellite is calculated from the orbital elements using the standard relations

$$\left.\begin{aligned} \mu = \frac{360^\circ}{P}, \quad M = \mu (t - T) = E - e \sin E, \\ r = a (1 - e \cos E), \quad \operatorname{tg} \frac{v}{2} = \sqrt{\frac{1+e}{1-e}} \operatorname{tg} \frac{E}{2}; \\ \operatorname{tg} (\theta - \Omega) = \operatorname{tg} (v + \omega) \cos i, \quad \varrho = r \cos (v + \omega) \sec (\theta - \Omega). \end{aligned}\right\} \qquad (8)$$

DIFFERENTIAL CORRECTIONS. The elements can be adjusted using new observations to obtain differential corrections. We have just written the expressions for the calculation of θ and ϱ from the orbital elements, which show that the position angle is a function of all the elements except the semimajor axis, and the distance is independent of the angle, i.e.,

$$\theta = f (\Omega, \omega, i, e, P, T),$$

$$\varrho = \varphi (a, \omega, i, e, P, T).$$

If $\Delta\theta$ and $\Delta\varrho$ are the deviations of the calculated points of the orbit from the results of new observations, and $\Delta\Omega$, Δa, $\Delta\omega$, etc., are the sought corrections, we can write

$$\Delta\theta = \frac{\partial f}{\partial \Omega} \Delta\Omega + \frac{\partial f}{\partial \omega} \Delta\omega + \ldots,$$

$$\Delta\varrho = \frac{\partial \varphi}{\partial a} \Delta a + \frac{\partial \varphi}{\partial \omega} \Delta\omega + \ldots$$

The expressions for the partial derivatives are obtained by differentiating Eqs. (8), which give

$$\frac{\partial f}{\partial \Omega} = +1, \quad \frac{\partial f}{\partial a} = \frac{\varrho}{a}, \quad \frac{\partial f}{\partial \omega} = \cos^2 (\theta - \Omega) \sec^2 (v + \omega) \cos i, \text{ etc.}$$

The coefficients are different for different observations, and the sought corrections are obtained by the least squares technique.

OTHER METHODS OF ORBIT DETERMINATION. Numerous alternative methods are available for the determination of orbital elements of visual binaries. Some of them are slight modifications of the above techniques. For example, the method of Anroto and Stuart, as was demonstrated by Kulikovskii, is in fact Mlodzeevskii's method /6/. Some methods are further improvements and elaborations of older techniques. Special methods have been proposed for cases when a short arc of a long-period binary is available from observations, such as Rabe's method /7/ or the method of Guntzel-Lingner /3/. All these methods, however, fail if the observations cover too short a period of time, whereas they are all equally satisfactory if the observations cover the entire orbital period. In some cases, special reduction techniques are used. For example, if the inclination i is close to 90°, the apparent ellipse is nearly a straight line. The drawing gives only the angle Ω. The other elements can be found by an ingenuous technique, which is described in some detail by Aitken /8/.

In some cases the satellite is so weak that it is invisible. The orbital motion of the primary about the center of mass is detected relative to reference strs. The projection of the focus of the true orbit, i.e,, the projection of the center of mass, is not available in this case. This introduces additional difficulties in the determination of the elements. We have to find a certain point inside the apparent ellipse such that the projections of the radii-vectors from this point sweep out equal areas in equal times.

As we have mentioned in the preceding, the orbit of very few stars are known. In 1938, Finsen /9/ published a catalogue listing the elements of 196 stars. Of these, some 90 binaries have fairly certain orbit, for 65 binaries the elements are tentative, and the remaining orbits are uncertain. Baise's catalogue /10/, published in 1950, contains 253 binaries with known orbits. For numerous stars, however, two or more systems of elements by different authors are given. The number of known orbits has by now reached 400 and is steadily increasing. The elements sometimes show considerable divergence, especially for the long-period stars (e.g., a few decades for the orbital periods, several tens of degrees for the angles, a few angular seconds for the semimajor axes, and a few tenths for the eccentricity).

§14. Determination of masses of binaries. Sum of the masses. Dynamic parallaxes. Our information on stellar masses is entirely based on the study of binary stars. By Kepler's third law,

$$\mathfrak{M}_1 + \mathfrak{M}_2 = \frac{a^3}{P^2},$$

where $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are the mass of the primary and the satellite, in units of solar mass, a is the semimajor axis of the relative orbit in astronomical units, P is the orbital period in years. Since the observations give the semimajor axis in seconds of arc, we require the parallax p of the binary. Then

$$\mathfrak{M}_1 + \mathfrak{M}_2 = \left(\frac{a''}{p''}\right)^3 \frac{1}{P^2}. \qquad (9)$$

Of all the variables in Eq. (9), the least certain is the parallax, which unfortunately enters this equation to the power of three. The error in the stellar masses is therefore mainly determined by the uncertainty in stellar distances. A better policy is therefore to solve the inverse problem, namely to find the parallax assuming known masses:

$$p'' = \frac{a''}{(\mathfrak{M}_1 + \mathfrak{M}_2)^{1/3}} \frac{1}{P^{2/3}} .$$

It is a fact that stellar masses, especially those of the main-sequence stars, show a fairly small dispersion. If we assume that the mass of any star is equal to one solar mass, i.e., unity, we will nevertheless obtain a satisfactory value of the parallax. To find the statistical average distance of stars without using the orbital elements, we can take the equality

$$P = \frac{2\pi\varrho}{v} ,$$

where ϱ is the radius of the circular orbit, equal to the measured distance between the components of the pair; v is the arc that the satellite traverses in one year, found by dividing the entire observed path by the time of observation in years. Taking 2 for the sum of the stellar masses, we get

$$p = \frac{\varrho^{1/3} v^{2/3}}{4.3} .$$

The parallaxes calculated in this way for binary stars are known as dynamic parallaxes. They are significant in that they give the distances of the more distant stars, whose trigonometric parallaxes are unreliable. Some 3000 dynamic parallaxes are available at present.

Mass ratio. Astrometric observations give the mass ratio of the binary components, which together with the sum of the masses enables us to find the individual masses of the components.

Suppose the position of each component is measured relative to reference stars. Let L_1 and L_2 be the distances of the primary A and the satellite B from the center of mass at any time t. The rectangular coordinates of the stars and the center of mass are X_1, X_2, X_c and Y_1, Y_2, Y_c. Clearly,

$$\frac{\mathfrak{M}_1}{\mathfrak{M}_2} = \frac{L_2}{L_1} = \frac{X_c - X_2}{X_1 - X_c} = \frac{Y_c - Y_2}{Y_1 - Y_c} .$$

Hence

$$\frac{\mathfrak{M}_1 + \mathfrak{M}_2}{\mathfrak{M}_2} = \frac{X_1 - X_2}{X_1 - X_c} = \frac{Y_1 - Y_2}{Y_1 - Y_c} .$$

For the X coordinate we have

$$X_1 = X_c + (X_1 - X_2) \frac{\mathfrak{M}_2}{\mathfrak{M}_1 + \mathfrak{M}_2} .$$

At some other time t', different coordinates will be observed due to the effect of orbital motion, as well as proper motion and the parallax. For the time t' we may therefore write

$$X_1' = X_c + \mu_x (t' - t) + p h_x + (X_1' - X_2') \frac{\mathfrak{M}_2}{\mathfrak{M}_1 + \mathfrak{M}_2} .$$

This equation has four unknowns: X_c, μ_x, p and $\frac{\mathfrak{M}_2}{\mathfrak{M}_1+\mathfrak{M}_2}$. For the other coordinate we have a similar equation

$$Y_1' = Y_c + \mu_y (t - t') + p h_y + (Y_1' - Y_2') \frac{\mathfrak{M}_2}{\mathfrak{M}_1 + \mathfrak{M}_2},$$

where two unknowns are the same as in the first equation. Given a sufficient number of observations, we can find the sought quantities by the least squares method. To ensure a reliable determination of the mass ratio, the observations should be spread over most of the orbit, so as to allow the coefficients $(X_1 - X_2)$ and $(Y_1 - Y_2)$ to change between significant limits. As an example, Table 2 lists the masses of some binaries from Eggen's recent observations /11/.

TABLE 2

Star	Spectrum	M_{bol}	$\frac{\mathfrak{M}}{\mathfrak{M}_\odot}$
Sirius A	A1V	+1.13	2.53
Sirius B	WA	—	1.08
α Cen A	dG4	+4.30	1.07
α Cent B	dK5	+5.65	0.87
70 Oph A	K0V	+5.68	0.82
70 Oph B	—	+6.82	0.60
Krü 60 A	dM4	+9.29	0.28
Krü 60 B	dM6	+10.43	0.17

According to Eggen, the mass—luminosity relation for 35 visual binaries is

$$L = \mathfrak{M}^{3.1}$$

or in the logarithmic form

$$M_{bol} = 4^m.79 - 7^m.8 \lg \mathfrak{M},$$

where M_{bol} is the absolute bolometric magnitude of the star, $4^m.79$ is the absolute bolometric magnitude of the Sun.

§15. Dark companions. The concept of a dark companion is somewhat arbitrary. In a sense, it means that the companion is not luminous and is thus invisible. In fact, however, the dark companions remain invisible because their luminosity is much too low or they are too near the brighter primary. The dark companions may become visible as the observation techniques improve.

The detection of dark companions is a highly complex undertaking. Over 100 years ago, Bessel discovered certain fluctuations in the motion of Sirius and Procyon after reducing long series of meridian-transit observations. This led him to the conclusion that some invisible satellites caused perturbation in the motion of these stars. The amplitude of the deviations reached 4 seconds of arc, orbital periods of some 50 years were obtained, and the masses were comparable with the mass of the Sun. Soon after that the mysterious satellites were noticed through the large

refractors and thus reclassified as visible companions. They were found to be white dwarfs.

Much smaller fluctuations, reaching 0".6, were discovered for Ross 614. It took the 200-in. Mount Palomar reflector to photograph the predicted satellite. Its photovisual magnitude was found to be $14^m.8$, and mass a mere 0.08 of the Sun /12/. The mass of the primary was $0.14\odot$. Both these masses are the smallest of the currently known stellar masses. Before this determination, the smallest mass was that of the companion of Krüger 60, equal to $0.17\odot$ (see last row in Table 2).

Even smaller fluctuations, reaching 0".016, were noticed in 61 Cyg /13/. The mass of the dark companion which causes these perturbations was found to be $0.008\odot$, only eight times the mass of Jupiter. Simple calculations show that the radius of the orbit described by the Sun under the perturbing influence of Jupiter is 0".004 if viewed from the nearest star α Cen. This value lies at the threshold of modern astronomical measurements using photographic negatives. The search for dark companions with masses of the order of $0.01\odot$ is thus limited to stars not farther than 10 pc. There are some 200 known stars within this radius. Table 3 lists data on some of the known or suspected dark companions.

TABLE 3

Star	1900.0 α	1900.0 δ	m_v, spectrum	P, years	a, e	$\mathfrak{M}/\mathfrak{M}_\odot$, p
η Cas	$0^h43^m.0$	+57°17′	3.6—7.4 G0—K0	24	0".019 0.6	0.01 0".182
Ross 614	6 24 .3	−2 44	11.3—14.8 M6	16.5	0".306 0.36	0.08 0".250
Ci 1244	10 14 .2	+20 22	9.5 M5	26.5	0".11 0.6	0.03 0".211
Lalande 21185	10 57 .9	+36 38	7.5 M2	8.0	0".034 0.30	0.01 0".398
70 Oph	18 0 .4	+2 31	4.3—6.0 K0	17.0	0".015 —	0.01 0".188
61 Cyg	21 2 .4	+38 15	5.6—6.3 K5—K5	4.9	0".010 0.63	0.008 0".292

The fourth column of the table lists the apparent visual or photovisual magnitudes and the spectra of the visible components. The fifth and the sixth column give the orbital elements; the semimajor axis is that of the visible star, whose motion is perturbed by the dark companion. The seventh column gives the masses of the dark companions and the parallaxes. If we are dealing with a binary, it is not always clear which of the two components holds the dark companion. There is also a possibility that the observed perturbations are the combined effect of two dark companions, say. The mass of the dark companion for 61 Cyg was calculated assuming a single satellite, and since the two visible stars have equal masses there is no point in trying to decide to which of the two stars this satellite belongs.

The mass of the dark companion is calculated as follows. Let a_1 and a_2 be the semimajor axes of the absolute orbits of the visible star and its dark companion, $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are their masses. The common center of mass of the system is at the focus of the orbits. Then

$$\frac{a_2}{a_1} = \frac{\mathfrak{M}_1}{\mathfrak{M}_2}.$$

On the other hand, by Kepler's law

$$\frac{(a_1+a_2)^3}{P^2} = \mathfrak{M}_1 + \mathfrak{M}_2.$$

where a_1+a_2 is the semimajor axis of the relative orbit. Eliminating a_2 between these equations, we get

$$\frac{a_1''^3\left(1+\frac{\mathfrak{M}_1}{\mathfrak{M}_2}\right)^3}{P^2} = \mathfrak{M}_1 + \mathfrak{M}_2,$$

which gives $\mathfrak{M}_2$. Since this mass is generally small compared to the mass of the visible component, the problem can be solved by successive approximations using the equation

$$\mathfrak{M}_2 = \frac{a_1''^3}{p}\left(\frac{\mathfrak{M}_1+\mathfrak{M}_2}{P}\right)^{2/3}. \tag{10}$$

Here p is the parallax. To first approximation we may take $\mathfrak{M}_2 = 0$ in the right-hand side of (10).

The determination of the mass of the dark companion is further complicated by the fact that the measurements of the visible component strictly speaking refer to the common photometric center, if the invisible companion has some luminosity. Let l_1 and l_2 be the distances of the geometrical centers of the star and the dark companion from the photometric center, and I_1 and I_2 their respective luminosities. We may naturally write

$$\frac{l_1}{l_2} = \frac{I_2}{I_1} \quad \text{or} \quad \frac{l_1}{l_1+l_2} = \frac{I_2}{I_1+I_2} = \beta.$$

If L_1 and L_2 are the distances of the star and the companion from the center of mass, we have

$$\frac{L_1}{L_1+L_2} = \frac{\mathfrak{M}_2}{\mathfrak{M}_1+\mathfrak{M}_2} = B,$$

and clearly $l_1+l_2 = L_1+L_2$. Subtracting the first equation from the second, we get

$$\frac{L_1-l_1}{L_1+L_2} = B-\beta.$$

where L_1-l_1 is the distance of the photometric center from the center of mass. Clearly the orbit described by the photometric center and its semimajor axis α are smaller than the true orbit of the visible component and its semimajor axis a. Hence, the mass ratio B is greater than that obtained from observations, $B-\beta$. Returning to Eq. (10), we can find the true mass of the dark companion $\mathfrak{M}_2$ by calculating the ratio of the measured and the true semimajor axes of the orbit. We have

$$\beta = \frac{I_2}{I_1+I_2} = \frac{1}{1+\frac{I_1}{I_2}} = \frac{1}{1+2.5^{\Delta m}}.$$

Taking $I \sim \mathfrak{M}^3$, we get

$$B = \frac{1}{1+\frac{\mathfrak{M}_1}{\mathfrak{M}_2}} = \frac{1}{1+\left(\frac{I_1}{I_2}\right)^{1/3}} = \frac{1}{1+2.5^{\Delta m/3}} = \frac{a_1}{a_1+a_2}.$$

Then

$$B-\beta=\frac{1}{1+2.5^{\Delta m/3}}-\frac{1}{1+2.5^{\Delta m}}=\frac{\alpha}{a_1+a_2},$$

whence

$$\frac{B}{B-\beta}=\frac{a_1}{\alpha}=\frac{1}{1-\frac{1+2.5^{\Delta m/3}}{1+2.5^{\Delta m}}}. \qquad (11)$$

When $\Delta m = 0$, i.e., when both stars are of equal luminosity, the photometric center coincides with the center of mass, and its measurements do not reveal any orbital motion. For $\Delta m = 2$, we see from (10) and (11) that the mass of the dark companion should be increased by a factor of 1.6. For $\Delta m = 4$, it should be multiplied only by 1.1.

Eq. (11) was derived assuming a cubic mass—luminosity relation, which is however not always true. Moreover, the difference Δm should be taken in bolometric magnitudes.* It should be kept in mind, however, that the calculation of the factor β is in principle highly approximate, since the luminosities of the dark companions are not known and for very close visible components the determination of Δm and the measurement of the photometric center are unreliable. On the average, Eq. (11) shows a good fit with observations, as we see, e.g., from van de Kamp's data /14/.

§16. Wide pairs. The angular separation criterion adopted by Aitken in his definition of double stars is not satisfied for numerous faint stars, although the physical bond between them is unquestionable. A striking example of such distant satellites is provided by the 11 mag. star discovered by Innes in 1915; Innes assigned it to the α Cen binary from considerations of proper motions. Parallax measurements of the weak companion confirmed the existence of a bond between these stars. Table 4 lists some data on the triple system α Cen.

TABLE 4

Star	1900.0 α	1900.0 δ	m_v	Sp	μ_α	μ_δ	p
A	$14^h32^m.8$	−60°25′	0.3	G0	−3″.61	+0″.70	+0″.751
B	14 32 .8	−60 25	1.7	K5			
C	14 22 .8	−62 15	11.0	M	−3 .75	+0 .87	+0 .762

The orbital elements of component B relative to A are the following:

$P = 80$ years $\quad \omega = 52°.1$
$T = 1955.8 \quad \Omega = 25°.4$
$e = 0.52 \quad i = +79°.2$
$a = 17″.66 = 23.2$ a.u.

The mass of each star is one solar mass. Given the parallax and the angular distance of the weak satellite from the center of inertia of the two

* Finally the equality $I_1 : I_2 = I_2 : I_1$ should be replaced by $I_1 : I_2 = D(I_2) : D(I_1)$, where D is the photographic density of the image, with allowance for the sensitometric properties of the plate.

bright stars, we find 10,500 a.u. for its linear distance. Taking 11,000 a.u. as the approximate radius of the circular orbit of the weak companion, we obtain 730,000 years for its orbital period and a mean annual orbital motion of 0".07 or about 0.3 km/sec. The difference in the proper motions of AB and C is three times this figure, but the order of magnitude is correct. The satellite orbit is possibly not circular. The spatial velocity of α Cen is 32 km/sec. The orbital motion of the distant companion is thus much less than the spatial velocity of the stars. If such a system were distant 100 pc from the Sun, the angular radius of the orbit of the weak companion would be about 1 minute of arc, its orbital motion would be a few thousandths of a second, and the proper motion would be noticeable, 0".04.

An interesting example of a wide pair is provided by the star BD + + 4°4048 ($\alpha = 19^h12^m.0$, $\delta = +5°1'$, 1900.0), for which van Biesbroeck found a weak companion at a distance of 74" with a common proper motion of 1".4. The measured parallax of the weak companion was found to be equal to the parallax of the primary (0".168 ± 0".004). Both stars are of spectral type M, apparent visual magnitudes $9^m.1$ and $17^m.9$. The absolute magnitude of the companion is thus $+19^m.0$, which is the highest absolute stellar magnitude currently known. The separation of 74" is 440 a.u. in linear dimensions (ignoring the projection effect). The formal orbital period is 9000 years, and the annual orbital displacement is 0".05.

It follows that common proper motion of two stars with fairly large mutual separation and inconspicuous orbital motion is often a sufficient criterion of a physical bond between the two. Since the concept of a wide pair is not strictly defined (we cannot establish the exact limits for distances between the components), a better term would be double stars with common (in magnitude and direction) proper motion. For example, there is a pair of stars in Ursa Major (known as ι and 10) with a separation of 6° which have identical parallaxes and very close proper motions (see Table 5).

TABLE 5

Star	m_v	1950 0		Sp	μ	φ	p	r, km/sec
		α	δ					
ι	3.1	$8^h55^m.8$	+48°14′	A5	0".500	240°.2	0".066	+26
10	4.1	8 57 .6	+42 0	F5	0 .502	238 .8	0 .070	+13

Each of these stars has a near companion, and the companion of ι UMa is itself a close binary. The linear separation between ι and 10 is 1.4 pc, which is approximately the mean interstellar distance in the Sun's neighborhood. On the other hand, there are some very close binaries with common proper motions whose orbital motion has not been detected, although there is undoubtedly a physical bond between them. Table 6 lists some double and triple stars with common proper motions observed in Kapteyn's areas.

Stars in areas 15 and 68 are typical examples of triple systems where two stars form a close binary and the third lies farther off. Star No. 137 is a very close pair whose components merge into a single extended image in photographs. The separation of the components is about 3". The stars listed in the table in areas 29 and 35 are not numbered in Kapteyn's catalogue.

TABLE 6

Area	Star	m_{ph}	ρ"	μ_x (0".001)	μ_y (0".001)	Spectrum
15	175	11.6		−71	+81	G8
	176	12.3	24	−68	+94	—
	86	11.4	762	−69	+96	G7
18	508	12.9		+54	+52	G—K
			358			
	852	12.2		+50	+50	G1
29	—	7.2		−76	−93	F0
			54			
	—	13.2		−59	−98	—
35	—	17.0		−90	+122	
			19			
	—	17.0		−102	+118	
68	137 bin.	11.3		+35	−24	F8
			305			
	138	13.8		+32	−28	G5

In the great majority of cases neither the distances of stars nor their spatial velocities are known. Therefore, two stars with common proper motions observed in one direction may in fact lie at widely different distances from the Sun and have completely different spatial motions. Let us calculate the probability that such pairs are mere optical binaries.

Let some area Σ (e.g., on a photographic plate) covering a certain number of square degrees on the sky contain N stars distributed at random. Let us take a small area σ and calculate the probability of finding n stars in that area. The probability of finding one star inside the area σ is σ/Σ. For any combinations of n stars out of the total of N, the probability of finding n stars in the area σ is

$$P_n = C_N^n \left(\frac{\sigma}{\Sigma}\right)^n \left(1 - \frac{\sigma}{\Sigma}\right)^{N-n}.$$

The last factor indicates that $N - n$ stars must remain outside the area σ. As σ is small, the last term is close to unity and can be dropped.

The sought probability can also be obtained from Poisson's formula

$$P_n = \frac{\left(\frac{\sigma}{\Sigma} N\right)^n}{n!} e^{-\frac{\sigma}{\Sigma} N},$$

which has been tabulated for various values of $\frac{\sigma}{\Sigma} N$ and n (see, e.g., Boev, G.P., Teoriya veroyatnostei (Probability Theory). 1950).

For direct comparison with observational data we require not the probability but the expectation value of the number of binaries in the entire area Σ. To this end we should multiply the probability P_n by the number of subareas σ, which is equal to Σ/σ. The expectation value is thus

$$P_n \frac{\Sigma}{\sigma} = C_N^n \left(\frac{\sigma}{\Sigma}\right)^{n-1}. \qquad (12)$$

The proper motions of stars may be represented by vectors originating at a common center and terminating on some area S. This is the area of the ring between circles of radii μ_1 and μ_2. The probability that the components of a double or a multiple star will have equal (in magnitude and

direction) proper motions with a probable error of $\varepsilon\sqrt{n}$ is clearly $\left(\frac{s}{S}\right)^n$, where s is the ares of a circle of radius $\varepsilon\sqrt{n}$. Here ε is the probable error of one proper motion. The expectation value of this event is $\left(\frac{s}{S}\right)^{n-1}$.

The expectation of the number of double and multiple stars whose components lie inside the circle σ and whose proper motions differ in magnitude and direction at most by $\varepsilon\sqrt{n}$ is thus

$$\text{E. V.} = C_N^n \left(\frac{\sigma}{\Sigma}\right)^{n-1} \left(\frac{s}{S}\right)^{n-1}. \qquad (13)$$

The distribution of proper motions is assumed to be isotropic (all directions being equiprobable). If we wish to take into consideration the preferred directions of proper motions associated with parallactic displacement and Kapteyn's streams, we require the distribution as a function of the position angle φ. In practice, this distribution can be found by a direct count of the points — the ends of the vectors μ in Kapteyn's diagrams. Smoothing the results of these counts over different position angles, we obtain the ratio s/S as a function of φ. This ratio is found to change by no more than a factor of 4. The change is noticeable mainly for large proper motions.

Consider the following example. Let a Kapteyn area of 1 sq. degree = = 3600 sq. minutes contain 30 stars with proper motions

$$0''.030 \geqslant \mu \geqslant 0''.015.$$

The area of the circle σ is 12 sq. min. The probable error of each proper motion is $\varepsilon = \pm 0.003$.

For two stars the radius of the error circle in this case is $\sqrt{2}.0''.003 \approx \approx 0''.004$. The ratio of the area s of the error circle to the area S of the ring between the circles of radii $\mu_1 = 0''.030$ and $\mu_2 = 0''.015$ is

$$\frac{s}{S} = \frac{1}{42}$$

(inhomogeneities in the distribution of μ are ignored). For

$$n = 30 \quad \text{and} \quad \frac{\sigma}{\Sigma} = \frac{1}{300}$$

the expectation value is 1/30. In practice, the number of these double stars is greater by one order of magnitude, which points to a significant physical bond between many of them.

The above expectation value, as we have noted, strictly refers to the directions to the stars and the projections of their motions on the tangent plane. However, just as proximity of two directions in space and similarity of velocity projections are little probable, the probability that two stars accidentally occupy near positions in space and accidentally have similar spatial velocities is negligible. This goes to show the high probability of a real physical bond between close stars with similar proper motions.

§17. Triple and multiple stars. Triple and multiple stars are also a fairly common phenomenon, and as the observation techniques improve it becomes clear that some of the binaries have in fact three and more components.

A good example is ζ UMa, a double star known from antiquity: Alcor, the companion of Mizar, is a 4 mag. star visible with a naked eye at an angular distance of 11' from the primary. In 1650 Riccioli resolved Mizar through a small telescope into two stars of 2 and 4 mag. with a separation of 14". At the end of the 19th century, the brighter component of Mizar was found to be a spectroscopic binary with a separation of merely 0".012 between the components. As the parallax is known (0".04), all the distances in the ζ UMa system can be calculated; these are in descending order 16,000, 300 and 0.3 a. u. Another example is the sextuple star α Gem, Castor. Two components were resolved at the beginning of the 18th century and W. Herschel first established its orbital motion; the orbital elements are known with fair certainty at present. The eccentricity is 0.4 and the semi-major axis is 5".8, which for a parallax of 0".07 is equal to 83 a. u. At the 19th century a distant satellite with common proper motion was found at a distance of 73", or 1000 a. u. At the end of the 19th century and during this century, each of the components was identified as a spectroscopic binary with periods of 2.9, 9.2, and 0.8 days for stars A, B, and C and orbital radii of about 0.02 a. u.

No less interesting is the star ζ Cancri, which is a quadruple star with the two components A and B separated by 0".9 and an orbital period of 60 years. The third component C is distant 7".3 from the first two; its motion is perturbed, and O. Struve back in 1875 attributed this perturbation to the existence of a dark companion D. The orbital period of AB and CD about the common center of mass is 1140 years, and the orbital period of D and C is 17 years with $a = 0''.24$. The mass of each of the four stars is approximately equal to the mass of the Sun, and since D is an invisible companion it must be a low-luminosity white dwarf.

Another well-known quadruple star, ε Lyr, has been recently studied anew /15/. The following tentative orbital elements were obtained for each pair (Table 7).

TABLE 7

Stars	P years	a	e	ω	i	Ω	T years
ε_1 (AB)	1166	2".8	0.2	166°	±138°	29°	1152.4
ε_2 (CD)	585	3 .0	0.5	92	±120	17	1644.5

The similarity in the elements i and Ω points to coplanar orbits, i.e., orbits lying in parallel planes. During the 100 years of observations, the separation $\rho = 208''.0$ and the position angle $\theta = 172°.3$ between the two pairs of stars in ε Lyr did not change much. If we assume that the common orbit is coplanar with the orbits of each pair, the spatial distance between the two pairs is found to be 13,600 a. u. Seeing that the average spectra of all the stars in ε Lyr are of type A3 and the sum of the masses is approximately 10 solar masses, we obtain an orbital period of about 300,000 years.

Triple and multiple systems of the greatest significance are those where, as in the above systems, a clear division into pairs of stars or pairs and solitary stars is possible. The orbital motion in these systems is elliptic

and the perturbation due to the distant companion is not very prominent. These systems are therefore highly stable and move in Keplerian orbits for indefintely long time. The situation is entirely different in some triple and multiple stars where the distances between the components are all comparable. These systems can be treated as an open cluster with very few stars. As the stars in a cluster encounter one another, some stars may acquire supercritical velocities and escape from the cluster, which gradually dissipates. Triple and multiple stars of the open cluster type or, to use Ambartsumyan's term, of the Trapezium type (after the multiple star θ'Ori, the Trapezium in Orion, whose four components form a trapezium) are thus inherently unstable formations which break up fairly rapidly. The small number of Trapezium-type multiple stars confirms this conclusion: these are young stars of early spectral classes. Wallenquist /16/ examined the 2771 triple and multiple systems in Aitken's and Innes's catalogues below 9 mag., and found only some 400 Trapezium-type systems. The selection criterion was the ratio of larger to smaller distance. All stars with this ratio greater than 3 are classified as ordinary multiple stars. Ambartsumyan /17/ has shown, however, that Trapezium-type systems may include optical trapeziums and even pseudo-trapeziums, i.e., ordinary multiple systems which appear as a trapezium only in projection onto the celestial sphere. While the optical systems are relatively few, the pseudo-trapeziums are much more numerous. According to Ambartsumyan, the probability that a normal multiple system is projected into pseudotrapezium is 0.09. Thus it seems that more than half of the 400 systems classified by Wallenquist as trapeziums are in fact pseudotrapeziums. Calculations further confirm that the percentage of real trapeziums among A—K stars is negligible, whereas it is much larger for O—B stars. A number of Trapezium-type systems are found in O clusters which in their turn are nuclei of stellar associations.

A typical example of a Trapezium-type multiple star is the Trapezium in Orion, where the separations of the four stars A, B, C, D are $AB = 8''.7$, $AC = 13''.1$, $AD = 21''.6$. Studies of the internal motions in this system, which is about 400 pc from the Sun, do not give at this stage any reliable data on the trajectories of the individual stars or any indication of a common expansion of the system. Ambartsumyan's catalogue of Trapezium-type systems lists 108 multiple stars with most of the components at mutual distances of from 0.02 to 0.2 pc. In individual trapeziums the separation may reach almost 1 pc. The number of components is up to 7.

§18. The statistics of double and multiple stars. We have commented above on the great frequency of double and multiple stars in the Galaxy. The exact ratio of the number of these stars to the number of solitary stars is hard to establish, since it depends on the conditions of observation and observational selection. The observational selection effect invariably distorts the statistical results, a point to be remembered in statistical analysis of general regularities. In the immediate neighborhood of the Sun the selection effect is less significant, since the observations supply a more comprehensive sample. Table 8 lists the results for three spherical volumes centered at the Sun with radii of 5, 10, and 20 pc, for which the observations give the total number of stars, and the number of solitary, double, and triple stars. The last row gives the corresponding data for stars up to 15 mag. with proper motions $\mu \geqslant 0''.015$, obtained for the 115 northern

Kapteyn areas using similarity in proper motions /18/. The radius of the sphere containing these stars is approximately 500 pc.

TABLE 8

Volume	Total stars	Solitary	Double	Triple	K_1	K_2
$p \geqslant 0''.2$	55	31	9	2	35%	22%
$p \geqslant 0$ .1	257	143	45	8	37	18
$p \geqslant 0$ .05	1076	745	134	21	21	16
$p \geqslant 0$ .002	7466	7087	149	27	2.5	18

The last two columns give the percentage of double plus triple stars relative to the number of solitary stars (K_1) and the percentage of triple stars relative to the number of double stars (K_2). Each pair and each triplet are counted as one. The percentage of double and triple systems is seen to decrease with increasing volume because of observational limitations; the percentage ratio of triple to double stars, on the other hand, remains virtually constant (18% on the average).

Numerous studies have shown that double stars are not a unique group and do not differ in any way from the field stars of the Galaxy. Their kinematic properties, distribution in space, galactic concentration, and spectral characteristics are exactly the same as of solitary stars of corresponding subsystems and types. Wide pairs do not differ from ordinary binaries in this respect. The components of binaries may be stars from different sequences of the Hertzsprung-Russell diagram, and this also applies to the wide pairs /19/. Combinations of main sequence stars with white dwarfs are also possible. W. Luyten compiled a list of 32 pairs of this type /20/. Among these there is one double star LDS 275 with white dwarfs for both its components ($\alpha_{1950.0} = 9^h35^m.0$, $\delta_{1950.0} = -37°07'$, $\mu = 0''.37$, $m_{pg} = 14.6$, $m'_{pg} = 15.0$, $\varrho = 3''.7$, $\theta = 37°$). According to Luyten's preliminary calculations, the period of this star is $P = 685$ years and the semimajor axis is $a = 5''.2$, which for a hypothetical parallax of $p = 0''.06$ gives $a = 87.5$a.u. The sum of the masses is $1.4\odot$.

The difference of the apparent magnitudes Δm for double stars is equal to the difference of absolute magnitudes ΔM. The luminosity of the components may differ by as much as 10 magnitudes and statistical analysis taking account of observational selection shows that this luminosity difference is equiprobable for all ΔM values. The relation of ΔM to differences of spectral type for main-sequence stars on the whole follows the well-known regular trend, according to which the weaker component corresponds to a later spectral type. However, as we have noted above, in some combinations this trend is not observed. In giant systems the satellite is generally of the earlier spectral type. In some cases the components of one spectral type may greatly differ in luminosities.

The frequency of distances from the primary, or more precisely the distribution of the semimajor axes of the orbits is more difficult to determine as it is greatly affected by lack of reliable data on stellar parallaxes, as well as by observational selection. Statistical reduction of observational data shows that the number of satellites decreases with increasing

separation between components. This refers to the distribution of satellites in projection onto the celestial sphere. If we consider the spatial distribution of satellites, combining the primaries at one point, we can introduce the concept of the spatial density of satellites σ, as a function of the distance r between the components. Statistical data give

$$\sigma \sim r^{-3}.$$

The number of satellites is thus inversely proportional to the distance from the primary. In other words, this is the frequency of the semimajor axes. According to Ambartsumyan /21/, the density σ in case of statistical equilibrium in the distribution of the semimajor axes is proportional to $r^{-3/2}$. Comparing these two distributions with observational data we conclude that the semimajor axes of binaries are definitely not in statistical equilibrium since in this case the number of binaries should increase with increasing distance ϱ.

Turning to the distribution of masses in binaries, we see that widely different mass combinations are equally probable, down to systems with planet-like companions, as we have seen above. This conclusion fits the wide range of possible ΔM values. Kuiper found /22/ that the percentage of binaries among stars with masses greater than the Sun's is higher than among low-mass stars. This is possibly due to observational selection, since stars of higher mass are generally stars with better pronounced orbital motion of the satellites.

Numerous analyses of various relations between the orbital elements of binaries failed to reveal any significant statistical correlation. The most remarkable period — eccentricity dependence is probably a result of observational selection. Observations sometimes show that the eccentricity increases with increasing period. Binaries with large orbital periods are more often detected for large eccentricities than for circular orbits, since fast motion near the periastron is easier to detect. In short-period stars, conversely, circular orbits are easier to detect, since in highly eccentric orbits the position of the satellite near the apastron drops out, its motion there being too slow to be detected from the relative displacement of the components. * No period — spectrum relation was observed for visual binaries.

The tables below give numerical data which support the above analysis. According to Wierzbinski /23/, whos analyzed about 400 visual binaries with known orbits, the dependence between period and eccentricity is as shown in Table 9 and 10.

TABLE 9

P, years	N	$\bar{e}$	$\bar{P}$, years
0—40	94	0.444	21.78
40—80	91	0.493	57.58
80—120	57	0.559	97.62
120—160	35	0.505	140.42
160—300	54	0.540	228.5
300—500	15	0.479	384.6
500—700	19	0.531	606.3
700—1100	10	0.603	814.3
1100—1550	3	0.374	1280.5

TABLE 10

e	N	$\bar{P}$	$\bar{e}$
0.00—0.20	48	175.00	0.114
0.20—0.40	80	108.91	0.285
0.40—0.60	109	148.41	0.491
0.60—0.80	84	211.16	0.696
0.80—1.00	57	146.86	0.869
0.00—1.00	378	157.14	0.502

* At the periastron the components are sometimes not separated.

The correlation coefficient between these elements was found to be $r = +0.07 \pm 0.05$ (mean error), which fails to support the dependence of P on e. The distribution of stars in Table 10 is fairly symmetric about the maximum at $e = 0.5$. Couteau /24/ gave a more detailed frequency of eccentricities for 400 binaries and compared the results with the empirical relation $f(e) = k(e - e^2)^{3/4}$ (Table 11).

TABLE 11

e	% obs.	% calc.
0.0—0.1	3.4	4.0
0.1—0.2	9.3	8.4
0.2—0.3	12.4	11.1
0.3—0.4	10.7	12.9
0.4—0.5	13.4	13.7
0.5—0.6	14.4	13.7
0.6—0.7	12.2	12.9
0.7—0.8	10.7	11.1
0.8—0.9	8.8	8.4
0.9—1.0	4.6	4.0

TABLE 12

Spectral type	N	e	$\overline{P}$, years
B	4	0.41	278.7
A	59	0.49	181.2
F	59	0.55	155.7
G	139	0.52	155.6
K	83	0.48	146.7
M	21	0.40	133.8

The frequency of spectral types in visual binaries with the corresponding eccentricities and periods are listed in Table 12. We see that there is no dependence whatsoever between these characteristics. The spectra are those of the primary component. The data for the geometrical elements i, Ω, ω, the eccentricity e, and the period P fail to reveal any relation either. The correlation coefficients are very small with errors of the same order of magnitude. The semimajor axis a and the eccentricity e are not correlated. Kepler's third law, however, lead to a certain correlation between a and P:

$$r = +0.51 \pm 0.04.$$

No correlation between i and a was observed (Table 13).

TABLE 13

a	N	$\overline{i}$
0".00— 1".00	286	52°.0
1 .00— 2 .00	52	52 .8
2 .00— 5 .00	27	49 .5
5 .00—10 .00	9	50 .1
>10".00	4	56 .2
$a<30"$	378	51 .9

In conclusion, let us briefly consider the question of coplanar orbits for double and multiple stars. The parallel spatial attitude of the orbits of binaries near the Sun was repeatedly investigated, always with negative results. The orbital planes are randomly oriented in space. Agekyan /25/ considered the coplanarity of orbits in multiple systems, but in his usage

coplanarity was defined as coincidence of orbital planes inside the system. The problem was solved statistically, by studying the mutual position of the components in multiple systems. Also, 59 stars with known orbital elements were used. Comparison of theoretical findings with observations provided Agekyan with partial confirmation of the coplanar nature of orbits in triple and quadruple systems.

REFERENCES TO CHAPTER II

1. JACKSON, J.—MN 115:2. 1955.
2. MULLER, P.—B.A. 14, fasc. 4, 257. 1949.
3. GÜNTZEL-LINGNER, U.—AN 281:79. 1953.
4. SMART, W. Textbook of spherical astronomy. 1963.
5. SHAIN, G.A. Kurs astrofiziki i zvezdnoi astronomii (A Course in Astrophysics and Stellar Astronomy), Part II.— GTTI. 1934.
6. KULIKOVSKII, P.G.— Astron. Zh., 26, No.4:249. 1949.
7. RABE, W.— AN 280:1. 1951.
8. AITKEN, R. The binary stars. 1935.
9. Union Obs. Circ., No.100:466. 1938.
10. BAISE, P.—J.O. 33, No.1—2. 1950.
11. EGGEN, O.— Astron. J. 61, No.9. 1956.
12. LIPPINCOT, S.— Astron. J. 60, No.10. 1955.
13. DEICH, A.N.— Bull. Pulkovo Obs., No.146. 1951; No.160. 1957; No.166. 1960.
14. Van de KAMP, P.—Astron. J. 59, No.11:447. 1954.
15. GÜNTZEL-LINGNER, U.— AN 283:73. 1956.
16. WALLENQUIST, A.— Uppsala Astr. Obs., Annalen 1:5. 1944.
17. AMBARTSUMYAN, V.A.— Bull. Byurakan Observ., XV. 1954.
18. DEICH, A.N.— Bull. Pulkovo Observ., No.168. 1961.
19. KULIKOVSKII, P.G.— Astron. Circular, No.171; 16. 1956.
20. LUYTEN, W.— Obs. Univ. Minnesota, 1956.
21. AMBARTSUMYAN, V.A.— Astron. Zh., 14, No.3:207. 1937.
22. KUIPER, G.P.— ApJ 95:212. 1942.
23. WIERZBINSKI, S.— Acta Astronomica 8, 2:78. 1958.
24. COUTEAU, P.— J.O.43, No.3, 1960.
25. AGEKYAN, T.A.— In: "Voprosy kosmogonii," III, p.63. 1954.

V.A. KRAT

Chapter IV

SPECTROSCOPIC BINARIES

§19. **Radial velocity curves.** Stars with variable radial velocity were discovered in 1889. Variation of radial velocity was first observed in the spectra of ζ UMa and β Per. Periodic shift or splitting of spectral lines are often the only clue of the binary nature of these stars. Among the spectroscopic binaries, only about one sixth show periodic line splitting. Whether the spectral lines of the secondary component are visible or not naturally depends on the luminosity ratio of the two stars. If the difference in magnitudes is over $0^m.7$, the lines of the secondary component remain mostly undistinguishable. It is difficult to draw a fixed boundary between spectroscopic binaries and visual binaries. The currently known spectroscopic binaries have periods ranging from a few hours (0.1084 day for γ UMi) to 15 years (ε Hya) and more (ε Aur and others).

The determination of orbits of spectroscopic binaries is fundamentally different from the determination of orbits of visual binaries. The main difference is that for a visual binary one mostly determines the orbit of the satellite relative to the primary component. The absolute orbit can be found only if the mass ratio of the two components is known, and the linear dimensions (in kilometers or astronomical units) can be obtained only if the parallax of the system is additionally available.

In the case of spectroscopic binaries, on the other hand, we determine the orbit of one or both components relative to the center of inertia of the system. Determination of the orbit diameter does not require knowledge of the parallax, but the result invariably gives the major axis multiplied by $\sin i$. The position of the line of nodes cannot be determined, although the time of passage through the nodes can be found without difficulty; for spectroscopic binaries ω is the angular distance of the periastron from the ascending node in the direction of orbital motion. At the ascending node the component star has maximum positive velocity.

Additional elements have also to be considered. These include the period P, expressed in mean solar days, which is generally the observed period and because of the finite velocity of light we have for the true period P_0

$$P_0 = \frac{P}{1+\frac{\gamma}{c}},$$

where γ is the radial velocity of the system's center of mass; the half-amplitudes K_1 and K_2 of the radial velocity curves of each of the components

($K_1 + K_2$ is the half-amplitude of the radial velocity curve of one component relative to the other); the semiaxes a_1 and a_2 of the component orbits relative to the center of inertia of the system; the inclination i defined as the angle between the orbital plane and the plane of the sky (spectrographic observations do not give the inclination).

Of the various methods for the determination of the orbital elements of spectroscopic binaries, the most convenient are those of R. Lehmann-Filhés and Irwin /1/. Other methods have been proposed by Schwarzschild, Russell, Zurhellen, King, and others. The Lehmann-Filhés method is the one most commonly used and corrections to tentative elements are obtained by the least squares technique.

The main part of the job is the construction of radial velocity curves and determination of the period. Formally, the orbital elements can be determined if the radial velocities are known for six different times. In practice, however, much more numerous observations are required, spread fairly uniformly over the different phases of the period, and it is from these observations that the velocity curve is constructed. If the period is very short (a few hours), it can be readily found after a single night of observations. Such instances, however, are very rare. The period generally must be found by successive approximations. The following method can be used for the determination of the period in the first approximation. The observations are plotted on graph paper. A convenient observation series is then chosen (fairly frequent observations are essential) and plotted on transparent tracing paper. The transparency is then moved in the horizontal direction across the main plot (taking care not to change the corresponding ordinates) until the best fit is achieved. In this position, the time covered by observations is clearly equal to the period or is a multiple of the period. Exact determination of the period requires different observation series at an interval of a few years. All the observations are reduced to a time covering a single revolution; this is done by subtracting multiples of the period from the later observation epochs. Deviations from this curve at points of maximum and minimum for the most separated epochs indicate in what direction and by what amount the period should be adjusted. A more exact treatment utilizes the middle part of the velocity curve, where the change in velocity with phase is the fastest. This approach, however, is practicable only if the successive correction is small; otherwise an error of as much as one half-period is possible.

The initial epoch of the curve is chosen somewhat arbitrarily; the best choice is the epoch when v_r is equal to the velocity of the center of inertia γ, and the difference $(v_r - \gamma)$ changes its sign.

If the variation amplitude of the radial velocity is small, the significance of the effect can be established, according to Schlesinger, by dividing all the observations into groups with v_r falling between v_r and $v_r + \Delta v_r$, where v_r is a whole number of km/sec, and Δv_r is 1 or 2 km/sec. The number of observations in each interval is counted: it is proportional to the frequency of occurrence of the corresponding v_r. The frequency curve obtained from observations is compared with the curve of random errors; this provides an indication of the likelihood that the star being considered is indeed a spectroscopic binary. Depending on the values of e and ω, the deviations from the error curve will show a definite systematic trend.

The shape of the velocity curve gives an indication of the orbital elements. Figure 24 shows three typical velocity curves and the corresponding orbits. The points a, b of the curve correspond to the points a and b of the orbit where the radial component of the orbital velocity vanishes. The points c and d correspond to nodes. The velocity curve corresponding to a circular orbit (A) is a sine curve with a, c, b, and d spaced at equal intervals. For an elliptic orbit (B and C), the radial velocity curves vary depending on e and ω ($\omega = 0°$ for B and $\omega = 270°$ for C).

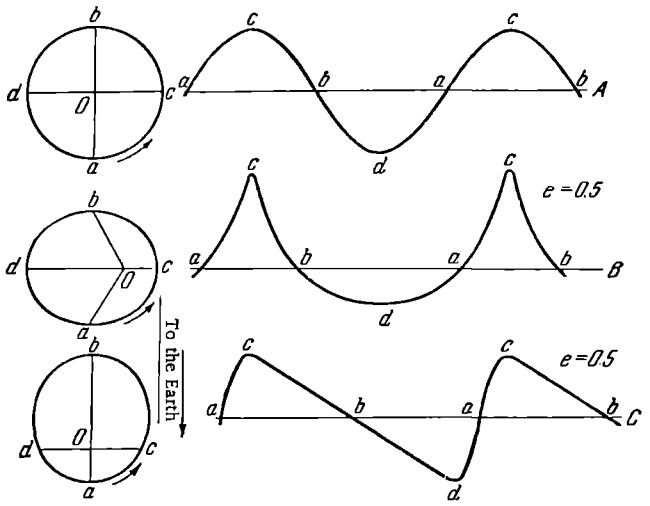


FIGURE 24

§20. **Fundamental equations for the determination of orbits. The Lehmann-Filhés method.** Let r and v be the radius-vector and the true anomaly of the point S in the orbit relative to the center of inertia of the system and z the projection of the radius-vector on the line of sight.

Then

$$z = r \sin i \sin (v + \omega). \tag{1}$$

The orbital component of the radial velocity v_r is equal to $\frac{dz}{dt}$ and is expressed in km/sec. Hence,

$$\frac{dz}{dt} = \sin i \sin (v + \omega) \frac{dr}{dt} + r \sin i \cos (v + \omega) \frac{dv}{dt}. \tag{2}$$

From the general equations of elliptical motion we have

$$r \frac{dv}{dt} = \frac{\mu a (1 + e \cos v)}{\sqrt{1 - e^2}} \qquad \left(\mu = \frac{2\pi}{P}\right), \tag{3}$$

$$\frac{dr}{dt} = \frac{\mu a \sin v}{\sqrt{1 - e^2}}, \tag{4}$$

whence

$$\frac{dz}{dt}=\frac{\mu a \sin i}{\sqrt{1-e^2}}[e\cos\omega+\cos(v+\omega)]. \tag{5}$$

This is the fundamental equation for the determination of orbits of spectroscopic binaries; v_r is related to $\frac{dz}{dt}$ by the equality

$$v_r=\frac{dz}{dt}+\gamma. \tag{6}$$

The line corresponding to $v_r=\gamma$ in radial velocity curves is called the γ axis.

It follows from Eq. (5) that for $v+\omega=0$ or 180°, $\frac{dz}{dt}$ reaches its maximum absolute value. Since ω is reckoned in the direction of motion from the ascending node to the periastron and v from the periastron in the same direction, $\left|\frac{dz}{dt}\right|$ clearly reaches its maximum value at the nodes. Let A and B be the values of $\frac{dz}{dt}$ at the nodes, reckoned from the γ axis, and K the half-amplitude; then

$$K=\frac{\mu a \sin i}{\sqrt{1-e^2}}, \tag{7}$$

$$A=K(1+e\cos\omega), \tag{8}$$

$$B=K(1-e\cos\omega), \tag{9}$$

$$\frac{A+B}{2}=K, \tag{10}$$

$$\frac{A-B}{2}=Ke\cos\omega, \qquad \frac{A-B}{A+B}=e\cos\omega. \tag{11}$$

Eq. (6) takes the form

$$v_r=\gamma+K[e\cos\omega+\cos(v+\omega)]=\gamma+\frac{A-B}{2}+\frac{A+B}{2}\cos(v+\omega). \tag{12}$$

Sometimes the ordinates are reckoned from the middle axis, $A=B$, and not from the γ axis. In this case

$$v_r=\gamma'+K\cos(v+\omega). \tag{13}$$

In Eq. (7) time is reckoned in days and K is expressed in km/sec. Reducing to a common time scale (seconds), we find

$$a\sin i=8.64\cdot 10^4\frac{KP}{2\pi}\sqrt{1-e^2}. \tag{14}$$

Eq. (14) can be written separately for the two component orbits (subscripting a and K with 1 and 2).

From Kepler's third law

$$\mathfrak{M}_1+\mathfrak{M}_2=\frac{(a_1+a_2)^3}{P^2}, \tag{15}$$

where the mass is measured in units of the Sun's mass ($\odot=1$), a_1+a_2 is expressed in astronomical units, and P in years. Since

$$\frac{a_1}{a_2}=\frac{\mathfrak{M}_2}{\mathfrak{M}_1},$$

we have

$$a_1 + a_2 = a_1 \frac{\mathfrak{M}_1 + \mathfrak{M}_2}{\mathfrak{M}_2}. \tag{16}$$

Inserting (16) in (15) and multiplying both sides by $\sin^3 i$, we get

$$(\mathfrak{M}_1 + \mathfrak{M}_2) \sin^3 i = \frac{1}{P^2} \left(\frac{\mathfrak{M}_1 + \mathfrak{M}_2}{\mathfrak{M}_2} \right)^3 a_1^3 \sin^3 i. \tag{17}$$

Inserting $a_1 \sin i$ from (14) and changing over to days (in P) and kilometers, we find the mass function

$$f_1 = \frac{\mathfrak{M}_2^3 \sin^3 i}{(\mathfrak{M}_1 + \mathfrak{M}_2)^2} = 1.038 \cdot 10^{-7} K_1^3 P (1 - e^2)^{3/2}. \tag{18}$$

If only the lines of one component are observed, we can only determine P, e, ω, f_1, and $a_1 \sin i$. If the lines of both components are observed, $\mathfrak{M}_1 \sin^3 i$ and $\mathfrak{M}_2 \sin^3 i$ can be obtained separately:

$$(a_1 + a_2) \sin i = 1.375 \cdot 10^4 (1 - e^2)^{1/2} (K_1 + K_2) P, \tag{19}$$

$$(\mathfrak{M}_1 + \mathfrak{M}_2) \sin^3 i = 1.038 \cdot 10^{-7} (1 - e^2)^{3/2} (K_1 + K_2)^3 P, \tag{20}$$

$$\left.\begin{aligned} \mathfrak{M}_1 \sin^3 i &= 1.038 \cdot 10^{-7} (1 - e^2)^{3/2} (K_1 + K_2)^2 K_2 P, \\ \mathfrak{M}_2 \sin^3 i &= 1.038 \cdot 10^{-7} (1 - e^2)^{3/2} (K_1 + K_2)^2 K_1 P \end{aligned}\right\} \tag{21}$$

and

$$\frac{\mathfrak{M}_1}{\mathfrak{M}_2} = \frac{K_2}{K_1}. \tag{22}$$

Using (16) we find $a_1 \sin i$ and $a_2 \sin i$ separately.

Let us now consider the determination of the orbital elements according to the Lehmann-Filhés method. The γ axis is first located on the radial velocity plot. Since it corresponds to the center of inertia of the system, the curve encloses equal areas on the two sides of the γ axis. Now, since the orbit is also a closed trajectory, the γ axis is clearly determined from the equality

$$\int_{t=0}^{t=P} \frac{dz}{dt}\, dt = 0. \tag{23}$$

In Figure 25 the areas DAC and CBD should be equal. The areas are readily measured with a planimeter or simply by counting the number of squares of the millimetric graph paper. Two or three trials are quite sufficient for placing the γ axis. After that the points with the ordinates A and B corresponding to maximum radial velocities (in absolute values) are found. These points are so chosen that the area AaD is equal to CbB and DaA is equal to BbD. Since C and D lie on the γ axis, we have from (5) and (11)

$$[\cos (v + \omega)]_{C,D} = -\frac{A - B}{A + B} = -e \cos \omega. \tag{24}$$

If v_1 is the true anomaly at point C (the star passes through this point on its way from ascending to descending node) and v_2 is the true anomaly at point D, $\sin (v_1 + \omega)$ is clearly positive and $\sin (v_2 + \omega)$ is negative.

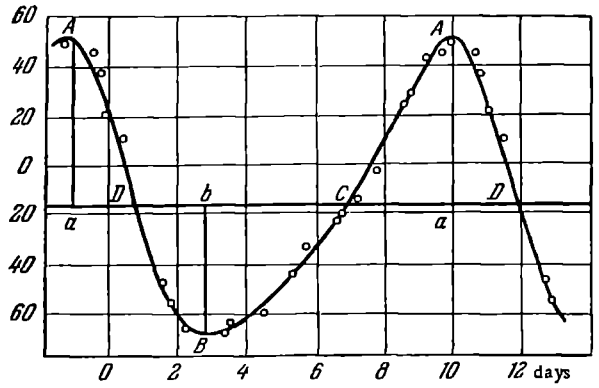


FIGURE 25

By (24)

$$\cos(v_1+\omega)=\cos(v_2+\omega)=-\frac{A-B}{A+B}. \tag{25}$$

Hence

$$\sin(v_1+\omega)=\frac{2\sqrt{AB}}{A+B},\quad \sin(v_2+\omega)=-\frac{2\sqrt{AB}}{A+B}. \tag{26}$$

Let the areas AaC and bBD be z_1 and z_2. If r_1 and r_2 are the radius vectors for the points C and D, we have

$$z_1=r_1\sin i\sin(v_1+\omega)=-z_2\frac{r_1}{r_2}.$$

Thus,

$$-\frac{z_1}{z_2}=\frac{r_1}{r_2}=\frac{1+e\cos v_2}{1+e\cos v_1}, \tag{27}$$

since

$$r=\frac{a(1-e^2)}{1+e\cos v}.$$

Writing $[(v+\omega)-\omega]$ for v, we obtain from (27)

$$-\frac{z_1}{z_2}=\frac{\sin(v_1+\omega)-e\sin\omega}{\sin(v_1+\omega)+e\sin\omega}, \tag{28}$$

whence

$$e\sin\omega=\frac{z_2+z_1}{z_2-z_1}\sin(v_1+\omega)=\frac{2\sqrt{AB}}{A+B}\frac{z_2+z_1}{z_2-z_1} \tag{29}$$

Eqs. (29) and (11) give e and ω if the areas z_1 and z_2 have been measured (z_2 is taken negative, for A and B their absolute values are used).

At the time of periastron passage $v=0$, and by (5) and (7) we have

$$\left(\frac{dz}{dt}\right)_{\text{per}}=K(1+e)\cos\omega. \tag{30}$$

There are clearly two points with this ordinate on the curve. For the points A and B, however, $(v+\omega)$ is 0 and 180°, respectively. The periastron is therefore determined without any ambiguity. The corresponding abscissa gives the phase of periastron passage. Adding this phase to the epoch selected for the beginning of the curve, we find the time T corresponding to the epoch of periastron passage.

From these elements we can easily find the ephemeris, which is then compared with observations to provide a check on the elements. The relations used are the following:

$$M = \mu(t-T) = E - e\sin E, \tag{31}$$

$$\operatorname{tg}\frac{1}{2}v = \sqrt{\frac{1+e}{1-e}}\operatorname{tg}\frac{1}{2}E, \tag{32}$$

$$v = \gamma + Ke\cos\omega + K_1\cos(v+\omega). \tag{33}$$

If e is less than 0.77, v as a function of M (the mean anomaly) is read directly from the tables published by the Allegheny Observatory /2/.

§21. The methods of Irwin, Schwarzschild, and others. The fundamental shortcoming of the Lehmann-Filhés method is felt when the radial velocity curve is fairly uncertain. The points A and B are sometimes difficult to compare with due precision, since here the derivative of the ordinate is zero. If only few observations are available, this may give wrong results for e and ω. Therefore Irwin /1/ forgoes the determination of the points A and B and concentrates instead on the points at which the ordinate of the velocity curve is $+0.7K$ and $-0.7K$ (reckoned from the middle axis) (Figure 26). This gives the points a, b, c, d, e, f, g, h, i (c and g correspond to the points A and B).

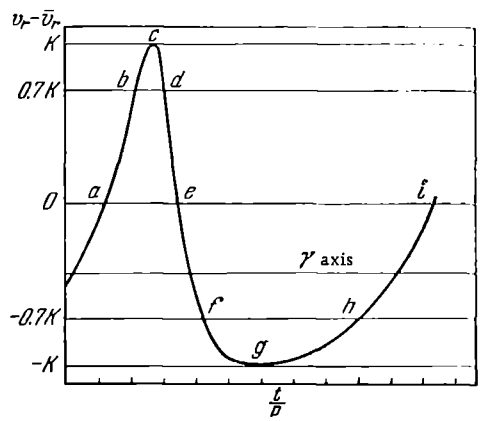


FIGURE 26

Then Irwin proceeds to calculate the ratios

$$u = \frac{t_e - t_a}{P} \tag{34}$$

(Irwin uses the letter h for u)

$$s_1 = \frac{t_b - t_a}{P}, \quad s_2 = \frac{t_c - t_d}{P}, \quad s_3 = \frac{t_f - t_e}{P}, \quad s_4 = \frac{t_i - t_h}{P}, \tag{35}$$

$$p_1 = \frac{t_c - t_b}{P}, \quad p_2 = \frac{t_d - t_c}{P}, \quad p_3 = \frac{t_g - t_f}{P}, \quad p_4 = \frac{t_h - t_f}{P}. \tag{36}$$

Each of the ratios u, s_1, s_2, s_3, s_4, p_1, p_2, p_3 and p_4 is only a function of two parameters, e and ω. The numerical values of these functions were tabulated by Irwin. Two of these ratios, say u and s_4, are sufficient for the determination of e and ω. Each of these functions is represented by a curve in the e, ω plane. The intersection of the curves $u(e, \omega)$ and $s_4(e, \omega)$ gives the first pair of values of e and ω. To obtain e and ω from curve intersection any of the following can be used: u, s_1, s_2, s_3, s_4, p_2, p_4; p_1 and p_3 are used to calculate the position of the points c and g on the velocity curve.

An original method was proposed by Schwarzschild /3/ and subsequently developed by Zurhellen /4/. This method also calls for the construction of the middle axis of the velocity curve.

An aribtary point is chosen on the curve and two points are marked on the horizontal axis at distances $\frac{P}{2}$ and $\frac{3P}{2}$ from it. A tracing paper is placed on top of the curve, the curve and the middie axis are drawn on the tracing paper and the points O, $\frac{P}{2}$ and $\frac{3P}{2}$ are marked. The replica is then moved 180° along the middle axis and turned over with its face down. The two velocity curves (the original and the replica) as a rule intersect at four points, two of which correspond to the periastron and the apastron. They are easily distinguished, since there is a time lag of $\frac{P}{2}$ between the two and they should lie on different branches of the velocity curve. That these points correspond to the periastron and the apastron, respectively, follows from the phase difference of 180° between the two and the difference of 180° in their true anomalies. In an ellipse these properties are characteristic only of the points which lie on the line of apsides. The periastron and the apastron are distinguished according to the following features: 1) the periastron lies on the steeper branch of the velocity curve, 2) the apastron lies on that side of the middle axis where the curve remains less time. If the eccentricity is close to zero, the periastron is clearly determined with considerable uncertainty.

After that ω is found. The position of the middle axis is given by

$$\bar{v}_r = \frac{v_A + v_B}{2} = \gamma + Ke\cos\omega. \tag{37}$$

The ordinate v_r relative to the middle axis is then given by

$$v_r = K\cos(v + \omega). \tag{38}$$

Since in the periastron and the apastron $v = 0$ and 180°, respectively, writing v_p and v_a for the corresponding ordinates, we get

$$\cos\omega = \frac{v_p}{K} \quad \text{and} \quad \cos\omega = -\frac{v_a}{K}. \tag{39}$$

The eccentricity is best found by Zurhellen's method. Other useful methods are the simple graphical techniques of Russell /5/ and King /6/.

§22. **Improving the elements by the least squares method.** The orbital elements obtained in this way should be improved by the least squares method /7/.

Differentiation of Eq. (6) gives

$$\delta v_r = \delta\gamma + [\cos(v+\omega) + e\cos\omega]\,\delta K + K\left[\cos\omega - \frac{\sin(v+\omega)\sin v}{1-e^2}(2+e\cos v)\right]\delta e - \\ - K[\sin(v+\omega) + e\sin\omega]\,\delta\omega + \sin(v+\omega)(1+e\cos v)^2\frac{K\mu}{(1-e^2)^{3/2}}\delta T - \\ - \sin(v+\omega)(1+e\cos v)^2\frac{K(t-T)}{(1-e^2)^{3/2}}\delta\mu. \qquad (40)$$

Here δv_r is the difference between the observed and the calculated velocities. This differential equation is preferably applied not to individual observations but to so-called "normal points," combining the observations at nearby phases into one. The period is determined with relatively high accuracy, so that the corresponding correction is dropped from (40). $\delta\gamma$, δK, δe, $\delta\omega$, and $\delta\mu$ — the corrections to the orbital elements — can be found by the least squares method. Eq. (40) should be written in the form

$$\delta v_r = \Gamma + \varkappa\cos u + \pi K\sin u + \alpha\varepsilon K\sin u + \beta\tau K\sin u + \beta K(t-T)\,m\sin u. \qquad (41)$$

Here, in addition to K we know the approximate values of

$$\alpha = 0.452\sin v\,(2+e\cos v), \qquad (42)$$

$$\beta = \frac{(1+e\cos v)^2}{(1+e)^2}, \qquad (43)$$

$$u = v+\omega. \qquad (44)$$

There are six sought unknowns:

$$\Gamma = \delta\gamma + e\cos\omega\,\delta K + K\cos\omega\,\delta e - Ke\sin\omega\,\delta\omega, \qquad (45)$$

$$\varkappa = \delta K, \qquad (46)$$

$$\pi = -\delta\omega, \qquad (47)$$

$$\varepsilon = -\frac{2.21}{1-e^2}\delta e, \qquad (48)$$

$$\tau = \mu\sqrt{\frac{1+e}{1-e}}\,\frac{1}{1-e}\,\delta T, \qquad (49)$$

$$m = -\sqrt{\frac{1+e}{1-e}}\,\frac{1}{1-e}\,\delta\mu. \qquad (50)$$

The values of α and β were tabulated by Schlesinger /7/.

If the spectra of both components are observed, the elements γ, e, T, P are clearly the same and there is a difference of 180° in ω. In practice, however, the determination of the individual orbits of the two components never gives identical results on account of the large observation errors. In this case it is therefore advisable to solve Eq. (40) for the two components simultaneously. This will increase the number of unknowns to seven, since δK is now replaced by two separate corrections δK_1 and δK_2. For the second component, we should replace ω by 180° + ω and write $-\delta\omega$ for $\delta\omega$. In this case equations of the form (41) cannot be used.

REFERENCES TO CHAPTER IV

1. IRWIN, I. B.— ApJ 116:218. 1952.
2. SCHLESINGER, F. and S. UDICK.— Publ. Allegheny Obs., 2, No. 17. 1912.
3. SCHWARZSCHILD, K.— AN 152:65, 1. 1900.
4. ZURHELLEN.— AN 175:245. 1907; 177:321. 1908.
5. RUSSELL, H. N.— ApJ 40:282. 1914.
6. KING, W. F.— ApJ 27:125. 1908.
7. SCHLESINGER, F.— Publ. Allegheny Obs., 1:33. 1908.

V.A. KRAT

Chapter V

PHOTOMETRIC BINARIES (ECLIPSING VARIABLES)

Some of the close binaries observed as spectroscopic binary systems are eclipsing variables. In these systems the angle ι is fairly close to 90°; the component stars therefore alternately occult one another in the course of their orbital motion. This effect produces a reduction in the brightness of the system; the binary behaves as a variable with a stable light variation curve. The first eclipsing variable was discovered in 1783 and studied by Goodricke. This was Algol, whose brightness variation apparently had been noted long before although its period had not been established. By now, over two thousand binaries of this type are known.

§23. **The light curve.** The shape of the light variation curve mainly depends on the individual brightness of the components relative to the total brightness of the system when not in eclipse, the relative dimensions of the stars, the distance between their centers of inertia, and the inclination (i). If the brightness of the weaker component, which is generally called the satellite, is not too small, the curve shows two minima: the principal (primary) minimum registered when the satellite occults (eclipses) the brighter (primary) component and a secondary minimum showing when the satellite is eclipsed by the primary. Observations show that the deepest minima correspond to a loss of no more than 4—6 magnitudes in brightness. The satellite is invariably a self-luminous body, a star. The shape of the light curve will greatly differ for stars which are luminous spheres and stars deformed by tidal forces. Tidal forces invariably produce a certain deformation of the components, but if the deformation is small, it does not have a significant effect on the light curve. In this case, the eclipsing binary is generally called an Algol variable.

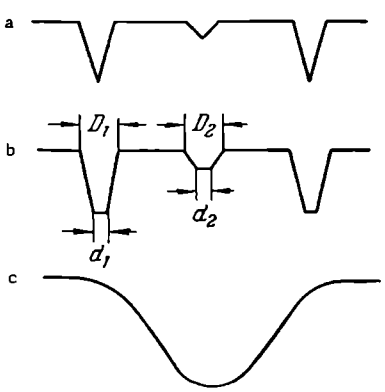


FIGURE 27

The light variation curve of an Algol variable depends on whether the eclipse in the primary minimum is partial, total, or annular. In the case of a partial eclipse (Figure 27a), the schematic light curve has two minima

with (roughly) constant brightness in between. In the case of a total eclipse (Figure 27b), a constant minimum brightness is observed for the duration of the time that the main star is totally eclipsed by the satellite. The case of an annular eclipse is not greatly different from the case of a total eclipse, but because of the nonuniform distribution of brightness over the visible part of the surface of the primary, the minimum brightness is not quite constant (Figure 27c)

If the tidal effects are very pronounced and the stars are viewed as two greatly flattened ellipsoids (β Lyr stars), the brightness of the system when not in eclipse does not remain constant: it varies between the primary and the secondary minimum in such a way that the two minima as if merge into a continuous formation (Figure 28). This is so because the disk areas* of the component stars in this case are smallest at the minima and largest in the quadratures, when the line of centers is perpendicular to the observer's line of sight. If, furthermore, the two components have virtually the same size and magnitude, both minima are similar in shape and have the same amplitude. Such light curves are often referred to as W UMa curves.

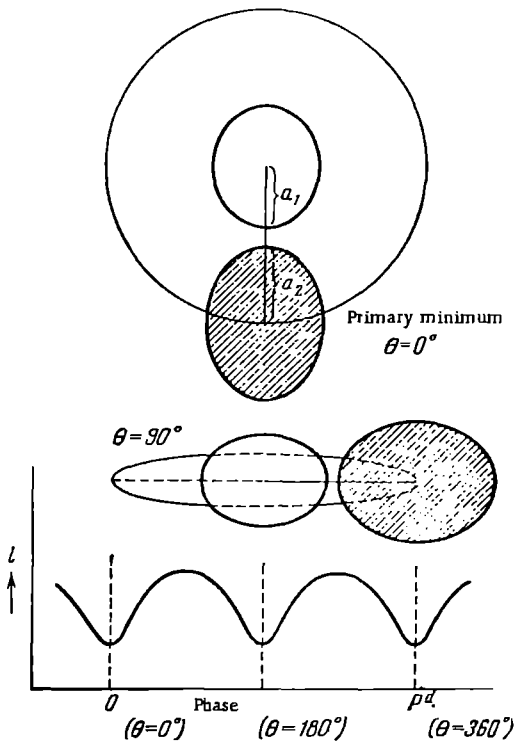


FIGURE 28

§24. Models of binaries. The observer's task is to determine the geometrical dimensions of the components, the geometry of their orbits, and

* The disk of a star is the visible surface of the star projected onto the plane of the sky; this is not necessarily a circle.

the values of the various physical factors affecting the shape of the light curve. In the final analysis, having found the numerical values of all the parameters which determine the light curve, we construct some theoretical light curve and compare it with the results of observations. To avoid various difficulties associated with side effects, which often cannot be allowed for with due accuracy, one generally uses essentially simplified and schematic models of the eclipse.

Three models of eclipsing binaries are in current use: 1) spherical stars, 2) a system of two similar ellipsoids, and 3) a sphere—ellipsoid system. The observer's task is to determine which of the three models best fits the system in question. It is only then that the observer can go on with his calculations and determine the so-called relative elements of the system (the linear dimensions are expressed in units of the semimajor axis):

Semimajor axis of the orbit $A = 1$	Inclination i
Eccentricity e	Period P
Periastron longitude ω	Epoch of the primary minimum T
Radius of the small star r_2	
Radius of the big star r_1	
Luminosity of the small star L_2	
Luminosity of the big star L_1	

The first model uses two spherical stars revolving in elliptical or circular orbit around a common center of inertia. In the second model we are dealing with two similar ellipsoids. In the third model one star is spherical and the other is an ellipsoid. The last case is most often applicable to systems with components of greatly different masses, when the tidal effects are substantial for one of the stars only.

For ellipsoidal stars r_1 and r_2 are the semimajor axes. In addition to the relative elements, we sometimes can measure a number of other constants characterizing various physical effects in the system: reflection of light from the atmosphere of the companion star, limb darkening, attenuation of light in the gas streams sometimes observed in close binaries, etc.

It is very difficult to decide a priori which of the three models best fits a particular binary. If the light curve does not provide a definite indication that the figures of the component stars are markedly non-spherical, the solution should start with the spherical model. In this case, if the radius of the large star (r_1) is found to be greater than 0.2 (the tidal deformation of this star is then quite substantial) and $L_1 \approx 0$ (the shape of the light curve between eclipses is independent of the figure of the large star), we should proceed to a sphere—ellipsoid model. If, on the other hand, the light curve shows definite signs of nonspherical shape of the components (the so-called ellipsoidal effect), the solution should start with the model of two similar ellipsoids, subsequently adopting the sphere—ellipsoid model if necessary. The latter is applicable only if the two stars greatly differ in their mass and dimensions, so that one of the stars is not markedly affected by tidal forces and can be regarded as a sphere.

If $A \sin i = (a_1 + a_2) \sin i$, $m_1 \sin^3 i$, and $m_2 \sin^3 i$ are available from spectroscopic observations, m_1 and m_2 can be found in units of the solar mass, and r_1 and r_2 can be expressed either in units of the Sun's radius or in kilometers. The absolute values of L_1 and L_2 (and hence the absolute magnitudes M_1 and M_2) can be found only if the parallax is known.

The determination of the relative elements of an eclipsing binary is fundamentally different from the determination of orbital elements of visual and spectroscopic binaries. In the latter case we are not dealing with elliptical motion of gravitating points whose exact law of motion is known; our aim is in fact to find the most probable values of the orbital elements. In eclipsing binaries, on the other hand, we can chart the exact motion of the components and our task is to determine a number of constants within the margin of error of our observations and then draw the light curve. The whole point is that the theoretical light curve must not give systematic deviations from the observational curve. This is not always feasible, since the simplified model ignores various effects which systematically influence the shape of the light curve. An error in the choice of the model may also sometimes produce biased deviation.

§25. The model of spherical stars. The U hypothesis. The first attempt to devise a method for the determination of the relative elements of an eclipsing variable (for spherical stars) is due to Blazhko /1/. Later Russell /2/ proposed a simple and convenient method, which is often referred to as the solution of the light curve. Since most of the current methods have grown out of Russell's technique, we will discuss it in some detail.

The first step is to change over from stellar magnitudes (if the light curve is given in stellar magnitudes) to the brightness l using the standard relation

$$\lg l = 0.4\,(m_0 - m), \tag{1}$$

where m_0 is the maximum magnitude corresponding to $L_1+L_2 = 1$; m_0 remains constant only if the component stars are sufficiently far from each other and the reflection of light from the stellar atmospheres does not markedly influence the light curve. A special table /1/ can be conveniently used for conversion of m and l.

The problem is solved under two extreme assumptions concerning the distribution of brightness over the star disk: 1) the disks are uniformly illuminated (the U hypothesis), and 2) the disks show total darkening toward the limb (the D hypothesis), the brightness B varying as a cosine of the angle γ between the observer's line of sight and the normal to the star's surface:

$$B = B\,(0)\cos\gamma, \tag{2}$$

$B\,(0)$ is the brightness at the center of the disk.

Let us first consider the solution using the U hypothesis. A circular orbit is assumed. We introduce a constant parameter $k = \frac{r_2}{r_1}$ and a function $\alpha\,(l)$ which is defined by the following expression for the minimum brightness (l_1), when the small star is eclipsed:

$$l_1 = 1 - \alpha L_2. \tag{3}$$

In the secondary minimum, when the satellite eclipses the same fraction of the surface of the primary (the $k^2\alpha$-th part of its disk), we have

$$l_2 = 1 - k^2\alpha L_1. \tag{4}$$

The parameter α is called the photometric phase of the eclipse. In a total eclipse $\alpha = 1$. In a partial eclipse $\alpha < 1$. The maximum value of α in this case is designated .

Using (3) and (4) and writing $l_1 = 1 - \lambda_1$, $l_2 = 1 - \lambda_2$ (λ_1 and λ_2 are the so-called light losses at the two minima) and $\alpha = \alpha_0$, we obtain the relation

$$1 - \lambda_1 \div \frac{1-\lambda_2}{k^2} = \alpha_0. \tag{5}$$

If the orbit is circular, the time of the middle of the primary minimum coincides with the time of the lower conjunction. The phase t of the light curve is generally reckoned from this instant. Let θ be the true longitude of the larger star:

$$\theta = \frac{360^\circ t}{P}. \tag{6}$$

All the values of t on the light curve should be converted to θ. The apparent distance between the centers of the two disks (δ), as we see from simple geometrical considerations (Figure 29), is related to the angles θ and i by the equality

$$\delta^2 = \sin^2\theta + \cos^2 i \cos^2\theta = \cos^2 i + \sin^2 i \sin^2\theta. \tag{7}$$

Since α depends only on k and $\varrho = \frac{\delta}{r_1}$, Eq. (7) can be written in the form

$$\varrho^2 = A + B\sin^2\theta, \tag{8}$$

where

$$A = \frac{\cos^2 i}{r_1^2}, \quad B = \frac{\sin^2 i}{r_1^2}. \tag{9}$$

The dependence of α on k and ϱ is best represented in tabular form /1/. Using this table, one finds ϱ as a function of k and α.

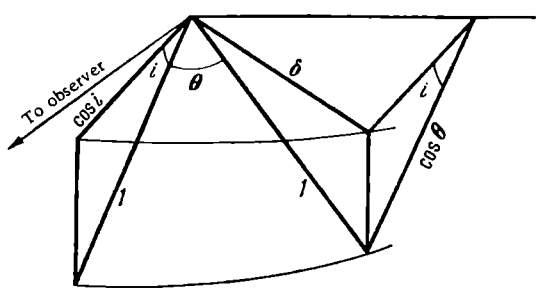


FIGURE 29

Russell /2/ was the first to demonstrate how for fixed θ the corresponding functions of k and α can be derived from the light curve. If the two branches of the primary maximum (the ascending and the descending one) are symmetric (it is only in this case that this procedure is applicable), the light curve is first "reflected" relative to the phase at the center of the

minimum, i.e., having calculated the phases of various points on the light curve taking the middle of the primary minimum as zero, we change the sign of the phases on the descending branch and plot them onto the ascending branch of the minimum. The average light curve is then drawn through the points of the ascending branch and the "reflected" points of the descending branch. This ensures a more reliable curve, which is appropriately called a "reflected" light curve. The minimum branch obtained in this way is generally divided into 10 parts at intervals of $0.1\,(1-\lambda_1)$ along the intensity axis and the corresponding angles θ_n are read off (Figure 30).

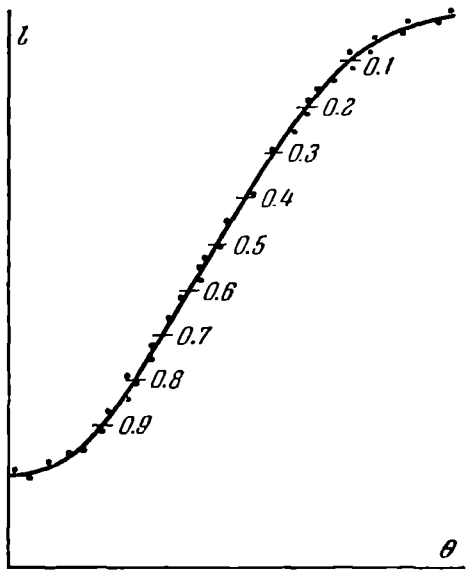


FIGURE 30

We write Eq. (8) for the time of the minimum ($\theta = 0$, $\varrho = \varrho_{1,0}$) and for some fixed $\theta = \theta_n$:

$$\left.\begin{aligned} \varrho_n^2 &= A + B\sin^2\theta_n, \\ \varrho_{1,0}^2 &= A. \end{aligned}\right\} \tag{10}$$

Subtracting $\varrho_{1,0}^2$ from ϱ_n^2, we get

$$\varrho_n^2 - \varrho_{1,0}^2 = B\sin^2\theta_n. \tag{11}$$

To eliminate the parameter B, we write Eq. (11) for $\alpha = 0.5\alpha_0$, i.e., for θ corresponding to half the depth of the primary minimum:

$$\varrho_{0.5}^2 - \varrho_{1,0}^2 = B\sin^2\theta_{0.5}. \tag{12}$$

The subscripts of ϱ and θ show what fraction α makes of α_0; in general $\alpha = n\alpha_0$. Dividing (11) through by (12), we get

$$\frac{\varrho_n^2 - \varrho_{1,0}^2}{\varrho_{0,5}^2 - \varrho_{1,0}^2} = \frac{\sin^2\theta_n}{\sin^2\theta_{0,5}}. \tag{13}$$

The left-hand side of Eq. (13) is a determinate function of k and α_0, since all the α are proper fractions of α_0. In other words, we may write

$$\chi(k,\ \alpha_0;\ n)=\frac{\sin^2\theta_n}{\sin^2\theta_{0.5}}. \tag{14}$$

Under the U hypothesis, all χ and $\sin^2\theta_n$ are related by the equalities

$$\chi(k,\ \alpha_0;\ n)=w_1(n)+w_2(n)\,\chi(k,\ \alpha_0;\ 0.25), \tag{15}$$

$$\sin^2\theta_n=w_1(n)+w_2(n)\sin^2\theta_{0.25}. \tag{16}$$

The functions $w_1(n)$ and $w_2(n)$ were tabulated by Russell. Using Eqs. (15) and (16) we can find the average values of $\sin^2\theta_{0.25}$ and $\chi(k, \alpha_0; 0.25)$ from all the $\sin^2\theta_n$ and $\chi(k, \alpha_0; n)$ corresponding to the points of the light curve with n from 0.1 to 0.9.

Now, to find α_0 and k, we can use the numerical value of $\chi(k, \alpha_0; 0.25)$ and Eq. (5). These two equations are best solved graphically: taking a certain k, the two equations are solved for α_0, and then taking a fixed α_0 we find k.

Since we do not know beforehand which of the two stars — the small or the large — is eclipsed in the primary minimum, Eq. (5) is represented by two curves corresponding to the two alternative propositions: (a) the smaller star is eclipsed in the primary minimum (the "large in front" hypothesis) and (b) the larger star is eclipsed in the primary minimum (the "small in front" hypothesis). Under the U hypothesis the pairs of k and α_0 for the function χ are the same in both cases, since the relation between k, α, and ϱ is derived for the case of "small in front" and α_0 is always calculated for that of the two minima for which this condition is observed. To obtain the maximum phase of the eclipse for the case of "large in front," α_0 should be multiplied by k^2.

From the pairs of values k, α_0 obtained for cases (a) and (b), two curves are plotted having a common point at $k = 1$. The same graph also shows the curve obtained from the k, α_0 values calculated from the numerical value of the function $\chi(k, \alpha_0; 0.25)$. The intersection of this curve with curve a or b gives the sought values of k and α_0. Figure 31 shows the intersection of the curve $\chi(k, \alpha_0; 0.25)$ and curve (5) for case (a). The solution was obtained for $k = 0.715$, $\alpha_0 = 0.845$.

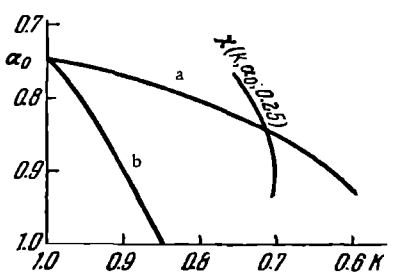


FIGURE 31

Sometimes the curves in this plot do not intersect. The effect of observational errors resulting in inaccurate numerical values for χ is often magnified by a poorly drawn light curve through average or so-called normal points based on several observations. In this case we should adjust the starting values of $\sin^2\theta_{0.25}$, $\sin^2\theta_{0.5}$, $1-\lambda_1$, and $1-\lambda_2$ within the accuracy of observations (within the dispersion of the normal points) so as to ensure intersection or at least contact of the curves in the plot. In practice, we first take the smallest $\Delta\chi$ (correction to χ) which ensures intersection and find the corresponding k and α_0, changing the initial data only after that.

In the final analysis, the actual method used to find the pair of values k and α_0 which in the observer's opinion satisfy Eq. (5) and the equation with $\chi(k, \alpha_0; 0.25)$ is irrelevant. The main thing is that the values of k and α_0 should ensure the best fit of the light curve. The theoretical curve is plotted using the values of θ_n calculated from Eq. (14)

$$\sin^2\theta_n = \chi(k, \alpha_0; n)\sin^2\theta_{0.5}.$$

The light curve plotted from θ_n should satisfy the following two conditions:

1) the sum of positive and negative deviations of normal points from the theoretical curve should be zero;

2) the differences $\theta_{obs} - \theta_{theor}$ (deviations) should be free from any bias.

To satisfy the first condition, it suffices to adjust $\sin^2\theta_{0.5}$. The second requirement, which is the best criterion of the accuracy of our solution, is more difficult to meet. k and α_0 should be somewhat varied until the best fit with the light curve is attained. In some cases we fail to obtain satisfactory fit for any k and α_0. In this case the U hypothesis is rejected and we have to use the D hypothesis.

In the case of a total eclipse $\alpha_0 = 1$, it remains to find k from the shape of the light curve at the minimum. Using Eq. (8), we construct the function $\psi_n(k)$:

$$\psi_n(k) = \frac{\varrho_n^2 - \varrho_{0.6}^2}{\varrho_{0.6}^2 - \varrho_{0.9}^2} = \frac{\sin^2\theta_n - \sin^2\theta_{0.6}}{\sin^2\theta_{0.6} - \sin^2\theta_{0.9}}. \qquad (17)$$

There are tables of $\psi_n(k)$ for various n. From each value of $\psi_n(k)$ we find the corresponding k, and then calculate the weighted average of all these k_n:

$$\bar{k} = \frac{\sum W_n k_n}{\sum W_n}. \qquad (18)$$

The weights W_n depend on the distribution of the observations along the light curve and on the derivatives $\left(\frac{\partial\psi}{\partial k}\right)_n$. Working equations can be found in specialized literature /3/. In practice, however, especially if the light curves are not particularly accurate, we can simply take the arithmetic average, using all k_n with $n < 0.6$. For $n > 0.6$, $\psi_n(k)$ become insensitive to variation in k, and therefore the corresponding W_n are so small that these k_n virtually make no contribution to the mean $\bar{k}$.

In case of a total eclipse, Eqs. (3) and (4) take the form

$$1 - \lambda_1 = L_2, \qquad (19)$$

$$1 - \lambda_2 = L_1 k^2, \qquad (20)$$

whence

$$k^2 = \frac{1 - \lambda_2}{\lambda_1}. \qquad (21)$$

For an annular eclipse the fraction in (21) should be inverted; the k obtained from Eq. (21) (generally designated k_0) in general does not coincide with $\bar{k}$. This is so because the U hypothesis does not provide a faithful representation of the brightness distribution over the stellar disks and also because the observational errors affect the depth of the minima and the

shape of the light curve. The reliability of k obtained from (21) depends on the number of normal points used to determine the constant brightness at the minima. The final value of k to be adopted should correspond to such λ_1 and λ_2 that the representation of the theoretical curve is the same everywhere. The representation for intermediate values of n is found by calculating the theoretical values of $\sin^2\theta_n$ from the equality

$$\sin^2\theta_n = (\sin^2\theta_{0.6} - \sin^2\theta_{0.9})\,\psi_n(k) + \sin^2\theta_{0.9},$$

where $\psi_n(k)$ is borrowed from tables for constant k. First k is taken as $\bar{k}$. If the observed curve deviates from the theoretical, the sum of the deviations of the normal points from the theoretical curve is made equal to zero by slightly varying $\sin^2\theta_{0.6}$ and $\sin^2\theta_{0.9}$ in such a way that the ratio $\frac{\sin^2\theta_n - \sin^2\theta_{0.6}}{\sin^2\theta_{0.6} - \sin^2\theta_{0.9}}$ remains fairly constant.

If the mean deviation of the normal points from the theoretical curve is such that within one half-amplitude of deviation Eq. (21) is satisfied for $k=\bar{k}$, we have found the solution. Otherwise, a second representation should be prepared using some average k, between $\bar{k}$ and k_0. The second approximation generally suffices to obtain a satisfactory representation of the two minima of the light curve.

Having determined k and α_n (partial eclipse) or k alone (total and annular eclipse), we can find A and B. For a total eclipse we have equations of the form (10), which putting $\varrho = 1 + kp(k, \alpha)$ ($p(k, \alpha)$ are also tabulated) can be written as

$$\left.\begin{aligned} [1 + kp(k, \alpha_0)]^2 &= A, \\ (1+k)^2 &= A + B\sin^2\theta_0. \end{aligned}\right\} \tag{22}$$

$\sin^2\theta_0$ is found from the curve representation:

$$\sin^2\theta_0 = \chi(k, \alpha_0; 0)\sin^2\theta_{0.5}. \tag{23}$$

By solving (22), we find A and B and hence the elements i and r_1 from the equalities

$$r_1^2 = \frac{1}{A+B}, \qquad \operatorname{tg}^2 i = \frac{B}{A}. \tag{24}$$

L_2 and L_1 are obtained from the obvious relations

$$L_2 = \frac{\lambda_1}{\alpha_0}, \qquad L_1 = \frac{\lambda_2}{k^2\alpha_0}. \tag{25}$$

The equality $L_2 + L_1 = 1$ provides a check of the theoretical results.

For a total eclipse, Eqs. (22) are replaced by

$$\left.\begin{aligned} (1-k)^2 &= A + B\sin^2\theta_{1.0}, \\ (1+k)^2 &= A + B\sin^2\theta_0. \end{aligned}\right\} \tag{26}$$

§26. The model of spherical stars. The D hypothesis. In the U hypothesis α was the ratio of the occulted disk area to the disk area of the small star, while in the D hypothesis α should be found by integrating $B(\gamma)$ over the occulted area. Now α is defined as the ratio of the light loss when the small

star is eclipsed to the total brightness of the small star. We call this ratio α'. If the larger star is eclipsed, we define the ratio α'' as the ratio of the light loss at a given time to the light loss at the second contact, when the annular phase of the eclipse may begin. We also introduce a function $Q(k, \alpha')$ which in a sense replaces the k of the U hypothesis. This function is defined by the condition that the ratio of the light loss of the eclipsed large star to its total brightness is $\alpha' Q(k, \alpha')$. At the second contact $Q(k, \alpha_0')$ reduces to $Q(k, 1)$. α_0' and α_0'' are thus related by

$$\alpha_0' Q(k, \alpha_0') = \alpha_0'' Q(k, 1). \tag{27}$$

The functions p, χ, and ψ in this case should be calculated separately for the two alternative hypotheses (a) "large in front" and (b) "small in front." In case (a) they depend on the parameter α_0', and in case (b) on α_0''. These functions were tabulated by Tsesevich /1/.

In case of a partial eclipse, the depths of the minima are expressed by the equations

$$\left.\begin{aligned} 1-\lambda_1 &= \alpha_0' L_1, \\ 1-\lambda_2 &= \alpha_0' Q(k, \alpha_0') L_2. \end{aligned}\right\} \tag{28}$$

Hence,

$$1-\lambda_1 + \frac{1-\lambda}{Q(k, \alpha_0')} = \alpha_0'. \tag{29}$$

On passing from the "large in front" to the "small in front" hypothesis, $1-\lambda_1$ and $1-\lambda_2$ are interchanged and then, using (27), we change over from α_0' to α_0''. In practice, Eqs. (29) are used to determine the function $Q(k, \alpha_0')$ from arbitrary given α_0' and after that k is found.

Eq. (27) is then applied to the pairs of α_0' and k values to find α_0''.

Subsequent stages of the solution are as in the U hypothesis, with the difference that the functions are represented by different graphs for the two alternatives of "large in front" and "small in front."

The general solution does not change (except for the use of the specific tables of $\psi_n(k)$ for the D hypothesis) when a total eclipse is considered. Note that the secondary minimum corresponds to an annular eclipse. In the annular phase $\alpha_0'' \geqslant 1$. The depth of the minima is given by

$$\left.\begin{aligned} 1-\lambda_1 &= L_2, \\ 1-\lambda_2 &= \alpha_0'' Q(k, 1) L_1. \end{aligned}\right\} \tag{30}$$

Thus, α_0'' is required for the representation of the secondary minimum and at least for the calculation of its theoretical depth. In practice one generally does not go into these calculations, since the D hypothesis need not apply to the secondary minimum either; this is indeed so if the secondary minimum is shallow and a marked phase effect is observed (see below, §35). The inaccuracy in the determination of $(1-\lambda_2)$ is mostly so large that there is not much point in even trying to represent the secondary minimum.

In some cases the primary minimum corresponds to a total eclipse, but the two minima are of comparable amplitude and we must consider the case of an annular eclipse. The solution for the annular eclipse is fundamental

when the primary minimum corresponds to an annular eclipse. The simplest solution for this case was proposed by Tsesevich /1/.

A typical D-hypothesis light curve for an annular eclipse has no sections of constant magnitude in the primary minimum for $\alpha'' > 1$. The time of the second contact and the corresponding angle are very difficult to establish. Since the secondary minimum corresponds to a total eclipse, we should have

$$1 - \lambda_2 = L_2 = 1 - L_1. \tag{31}$$

Let $1 - \lambda_0$ be the light loss at the second contact. Then by (30)

$$\lambda_2 = \frac{1 - \lambda_2}{Q(k, 1)}. \tag{32}$$

The angle θ at the second contact is designated θ''. Using the light curve, we assign a certain $1 - \lambda_0$ to any arbitrarily chosen θ'' and then from Eq. (32) find k. We thus obtain pairs of values k, θ''.

It now remains to find another relation between θ'' and k. This part of the problem can somewhat be simplified compared by Tsesevich's original method by reverting to Russell's technique. If $1 - \lambda_0$ is known, the interval $(1 - \lambda_0)$ can be divided into 10 equal parts as in the solution for the total eclipse (see Figure 30). We can then determine the value of Russell's function $\psi_n(k)$ for $n = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ (for the case of "large in front"), using this time the tables for the D hypothesis. From the numerical values of $\psi_n(k)$ we find k and its average over all the points. These average values $\bar{k}$ are obtained by taking various arbitrary $1 - \lambda_0$. We obtain pairs of values $1 - \lambda_0$ and $\bar{k}$ or, equivalently, θ'' and $\bar{k}$, using the shape of the light curve in the primary minimum (and not the depths of the minima). The two series of the $\bar{k}$, θ'' pairs are plotted on a graph. This gives two curves whose intersection point is the pair k, θ'' which solves our problem. α_0'' is clearly equal to the ratio $\frac{1-\lambda_1}{1-\lambda_0}$. The constants A and B, as in case of a total eclipse, are found from the equations

$$(1 - k)^2 = A + B \sin^2 \theta'',$$
$$(1 + k)^2 = A + B \sin^2 \theta_0.$$

The light curve is represented in the same way as in the U hypothesis.

§27. Intermediate cases of limb darkening. In fact neither the U nor the D hypothesis give the actual distribution of brightness over the disks of the components. Strictly speaking, because of the nonuniform distribution of the gravitational forces on the surface of a star deformed by tidal forces, different temperatures are observed at the poles, on the equator, and on the "tidal bulges." The distribution of brightness over the disk thus cannot be described as a function of one variable, the angle γ, as for the disk of a solitary star. Nevertheless, better approximations are generally obtained by taking the $B(\gamma)$ for the components of eclipsing binaries in the form

$$B(\gamma) = B(0)(1 - x' + x' \cos \gamma). \tag{33}$$

x' is the degree of limb darkening. For the U hypothesis $x' = 0$, and for the D hypothesis $x' = 1$. The larger the tidal deformation of the star, the greater is the deviation of the true B from the interpolation expression (33).

In this case we can do no better than find the solutions U or D. It is only for stars with small tidal effects (Algol variables) that the application of Eq. (33) is worthwhile and one should try to find x' from observations. Moreover, since for light curves of low accuracy the theoretical curves calculated from the U and D elements represent the light curve almost equally well, x' can be found only from highly accurate light curves, when the photometric phase of the eclipse is sufficiently large.

The methods for the determination of x' are dealt with by Semenova /4/. Without going into this question in any detail, it suffices to note that, once x' has been determined, we can find intermediate elements between the U and D solutions.

The light curve is extremely sensitive to x' in the case of an annular eclipse in the primary minimum. By taking x' other than unity, we clearly change only the form of the dependence between α'' and k. Thus, taking x_2' for the degree of darkening of the small star and x_1' for that of the large star, we replace α' and α'' by the functions $\alpha'^{x_2'}$ and $\alpha''^{x_2'}$. In case of total eclipse in the secondary minimum, we have $\alpha'^{x_2'} = 1$. Eq. (32) then takes the form

$$\lambda_2 = \frac{1-\lambda_0}{Q(k, 1, x_1')}. \tag{34}$$

If $1-\lambda_0$ in Eq. (34) is replaced by $1-\lambda_1$, we should change over in the denominator from $Q(k, 1, x_1')$ to $Q(k, \alpha_0''^{x_1'}, x_1')$:

$$\lambda_2 = \frac{1-\lambda_1}{Q(k, \alpha_0''^{x_1'}, x_1')} \tag{35}$$

According to Tsesevich /1/, $Q(k, \alpha_0''^{x_1'}, x_1')$ is expressed in the form

$$Q(k, \alpha_0''^{x_1'}, x_1') = \frac{\alpha_M''^{x_1'}}{\alpha_0''^{x_1'}} \frac{3-x_1'}{(3-3x_1')k^2+2x_1'[1-\sqrt{(1-k^2)^3}]}, \tag{36}$$

where $\alpha_M''^{x_1'}$ is the value of $\alpha''^{x_1'}$ corresponding to a central eclipse. For a given k, this parameter is written in the form

$$\alpha_M''^{x_1'} = 1 + A^{x_1'}(k), \tag{37}$$

where $A^{x_1'}(k)$ is a tabulated function of k. Writing $K(k, x')$ for the expression

$$\frac{(3-3x')k^2+2x'[1-\sqrt{(1-k^2)^3}]}{\alpha_M''^{x'}(3-x)},$$

we obtain from (35)

$$\alpha_0''^{x_1'} = \frac{1-\lambda_1}{\lambda_2} \frac{1}{K(k, x_1')}. \tag{38}$$

For the second contact $\alpha_0''^{x_1'} = 1$. $K(k, x_1')$ has also been tabulated. The subsequent stages are the same as in the D hypothesis.

Similar techniques were developed for all the different types of eclipses by Kopal.

Detailed tables for the calculation of elements by Russell's technique for different x' were also compiled by Merrill /5/.

§28. Modifications of Russell's method. The method of Fetlaar and the Kopal—Piotrovski method. Although Russell's technique remains highly useful to this day, various authors tried to simplify it and make it more reliable, so as to obtain immediately the best fit to the light curve. Some of the modifications of Russell's technique are in fact methods for graphic solution of the problem. The best known of these have been devised by Sitterly /6/ and Merrill /7/. In view of the complexity of the problem, graphic methods involve the use of numerous plots, which are not always of uniform accuracy. They are thus highly cumbersome and do not always ensure satisfactory accuracy of the elements. Graphic methods are therefore not in wide use. Two further modifications of Russell's technique were proposed by Sharbe and Krat /3/. Sharbe's method, further improved by Tsesevich /1/, is attractively simple and speedily gives the relative elements of the system. It is, however, not particularly reliable and generally yields a poor representation of the light curve. Krat's method was developed for the case of a partial eclipse. It differs from Russell's technique in that both the U and the D solutions are found by using all the θ_n of the "reflected" light curve (generally divided into 10 sections) in the primary minimum. The working functions are

$$\psi(k, n\alpha_0) = \frac{\sin^2\theta_n}{\sin^2\theta_{0.3}} = \frac{\varrho_n^2 - \varrho_{1.0}^2}{\varrho_{0.3}^2 - \varrho_{1.0}^2} \tag{39}$$

and

$$\chi(k, \alpha_0) = \frac{\sum_{0.4}^{0.9} \sin^2\theta_n - \frac{1}{2}[\sin^2\theta_{0.1} + \sin^2\theta_{0.2}]}{\sin^2\theta_{0.3}}. \tag{40}$$

The functions $\psi(k, n\alpha_0)$ and $\chi(k, \alpha_0)$ for the cases U and D have been tabulated in /8/ (for the D case, α_0 is replaced by α_0'). $\chi(k, \alpha_0)$ here replaces Russell's function $\chi(k, \alpha_0; 0.25)$; its main advantage is that it is always constructed using 10 points of the curve and is thus highly insensitive to changes in k and α_0. This function therefore generally gives a definite solution which rapidly leads to the best fit of the light curve. Experience shows that the Krat method elements cannot be substantially improved by the least squares method (see below).

When the light curve is represented using the function ψ, the "theoretical" values of $\sin^2\theta_n$ are calculated from

$$\sin^2\theta_n = \psi(k, n\alpha_0)\sin^2\theta_{0.3}. \tag{41}$$

Fetlaar's method /9/, improved by Piotrovski and Kopal, is fundamentally different from Russell's technique. The underlying idea is the application of successive approximations. Fetlaar writes Eq. (7) in the form

$$r_1^2(1 + kp)^2 = \cos^2 i + \sin^2 i \sin^2\theta \tag{42}$$

and transforms it to

$$\cos^2\theta \sin^2 i + r_1^2 + 2kpr_1^2 + p^2k^2r_1 = 1. \tag{43}$$

Let $\cos^2\theta = B$. The values of B are calculated, as before, for various $n(1 - \lambda_1)$ taking $n = 0.25, 0.30, 0.35, \ldots; 0.90, 0.95$. Fetlaar combines the

first six equations (for $n = 0.25, 0.30, 0.35, 0.40, 0.45, 0.50$) into a first group of equations. The second group comprises the equations with $n = 0.55$, 0.60, 0.65, 0.70, 0.75, and the third the equations with $n = 0.80, 0.85, 0.90$, 0.95. The arithmetic average of each group of equations is then calculated, and the system is thus reduced to three equations of the form

$$\left.\begin{aligned} \bar{B}_1 \sin^2 i + r_1^2 + 2\bar{p}_1 k r_1^2 + \bar{p}_1^2 k^2 r_1^2 = 1, \\ \bar{B}_2 \sin^2 i + r_1^2 + 2\bar{p}_2 k r_1^2 + \bar{p}_2^2 k^2 r_1^2 = 1, \\ \bar{B}_3 \sin^2 i + r_1^2 + 2\bar{p}_3 k r_1^2 + \bar{p}_3^2 k^2 r_1^2 = 1. \end{aligned}\right\} \tag{44}$$

Subtracting the second equation in (44) from the first and the third from the second, we get

$$\left.\begin{aligned} (\bar{B}_1 - \bar{B}_2) \sin^2 i + 2(\bar{p}_1 - \bar{p}_2) k r_1^2 + (\bar{p}_1^2 - \bar{p}_2^2) k^2 r_1^2 = 0, \\ (\bar{B}_3 - \bar{B}_2) \sin^2 i + 2(\bar{p}_3 - \bar{p}_2) k r_1^2 + (\bar{p}_3^2 - \bar{p}_2^2) k^2 r_1^2 = 0. \end{aligned}\right\} \tag{45}$$

Eliminating $\sin^2 i$ between these equations, we find

$$k = -2\frac{(\bar{B}_3 - \bar{B}_2)(\bar{p}_1 - \bar{p}_2) - (\bar{B}_1 - \bar{B}_2)(\bar{p}_3 - \bar{p}_2)}{(\bar{B}_3 - \bar{B}_2)(\bar{p}_1^2 - \bar{p}_2^2) - (\bar{B}_1 - \bar{B}_2)(\bar{p}_3^2 - \bar{p}_2^2)}. \tag{46}$$

Now k is calculated by successive approximations. The average values $\bar{B}_1$, $\bar{B}_2$, and $\bar{B}_3$ are obtained from the light curve. Here $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$, $\bar{p}_1^2$, $\bar{p}_2^2$, $\bar{p}_3^2$ are Fetlaar's coefficients. They have been tabulated /1/ as a function of k and α_0 for a partial eclipse and as a function of k alone for a total eclipse.

In case of a partial eclipse, we generally select α_0 and then use Eq. (46) to determine k by successive approximations, starting with $k = 1.0$, say. The successive approximations converge fairly rapidly, so that each α_0 corresponds to a certain k. After that, the pairs of k, α_0 obtained by this method are used with Eqs. (5) for the two cases of "large in front" and "small in front" to find k and α_0 graphically, as in Russell's technique.

In case of a total eclipse, the problem is simplified since Fetlaar's coefficients are independent of α_0 ($\alpha_0 = 1$) and the final value of k is obtained directly after several successive approximations using Eq. (46).

Fetlaar's method is seldom used in its original form. The main reason for this are its two fundamental shortcomings: 1) the calculations use only the lower part of the light curve ($n \geqslant 0.25$), whose shape is less sensitive to the system elements (it can be adequately fitted using different pairs of k and α_0) than the shape of the light curve in the minimum; 2) the method makes no provision for varying the initial values of $\sin^2 \theta_n$, although in view of the weak dependence of $\bar{p}$ and $\bar{p}_2$ on k and α_0 the errors in θ_n borrowed from the initial average light curve may lead to considerable errors in system elements. Fetlaar's elements therefore seldom ensure the best fit of the light curve.

The method of Kopal and Piotrovski /11/, based on Fetlaar's idea, is very popular. The main features of the method are best described for the case of a partial eclipse. Eq. (7) is written in the form

$$\cos^2 i + \sin^2 i \sin^2 \theta = \zeta^2 [1 + kp(k, \alpha)]^2. \tag{47}$$

For each normal point (i-th) we find α_i using the depth of the minimum $(1 - \lambda_1)$:

$$\alpha_i = \frac{1 - l_i}{1 - \lambda_1} \alpha_0 = n_i \alpha_0; \tag{48}$$

α_0 is related to $1-\lambda_1$ and $1-\lambda_2$ as before by Eq. (5). For $\theta = 0$, Eq. (47) takes the form

$$\cos^2 i = r_1^2 [1 - kp(k, \alpha_0)]^2. \tag{49}$$

Subtracting (49) from (47), we get

$$(p_i^2 - p_0^2)\, x + 2(p_i - p_0)\, y = \sin^2 \theta_i, \tag{50}$$

where

$$p_i = p(k, n_i \alpha_0), \qquad p_0 = p(k, \alpha_0),$$
$$x = r_1^2 k^2 \operatorname{cosec}^2 i, \qquad y = r_1^2 k \operatorname{cosec}^2 i.$$

After that, the solution proceeds by successive approximations, as in Fetlaar's technique, but the least squares method is also applied.

The depth of the minimum can be varied by introducing a correction to $1-\lambda_1$ directly in Eq. (50). Suppose that this correction shifts all the points on the plane with the coordinates n and $\sin^2\theta$ in the direction of the n axis by an amount

$$-n_i \frac{\Delta(1-\lambda_1)}{1-\lambda_1}. \tag{51}$$

This corresponds to a displacement

$$-n_i \frac{d \sin^2 \theta}{dn} \frac{\Delta(1-\lambda_1)}{1-\lambda_1}, \tag{52}$$

which should be introduced in the left-hand side of Eq. (50):

$$(p_i^2 - p_0^2)\, x + 2(p_i - p_0)\, y - n_i \frac{d \sin^2 \theta}{dn} \frac{\Delta(1-\lambda_1)}{1-\lambda_1} = \sin^2 \theta_i. \tag{53}$$

Now,

$$\frac{d \sin^2 \theta}{dn} = \frac{d \sin^2 \theta}{d\alpha} \frac{d\alpha}{dn} = 2y\alpha_0 (1 + kp) \frac{\partial p}{\partial \alpha}.$$

The error in the right-hand sides of Eqs. (50) is proportional to

$$\frac{d \sin^2 \theta}{dn} \Delta n. \tag{54}$$

Let w_i be the weight of each equation; $\sqrt{w_i}$ is inversely proportional to (54). We have

$$\frac{d \sin^2 \theta}{dn} \Delta n = \frac{d \sin^2 \theta}{dn} \frac{dn}{dl} \frac{dl}{dm} \Delta m = -\frac{d \sin^2 \theta}{dn} \frac{1}{1-\lambda_1} l \Delta m,$$

where m is the stellar magnitude. Assuming the errrors Δm to be the same for all the points of the light curve, we may take for $\sqrt{w_i}$

$$-\frac{1-\lambda_1}{l_i} \frac{dn}{d \sin^2 \theta}, \tag{55}$$

and Eq. (53) takes the form

$$\sqrt{w_i}\,(p_i^2 - p_0^2)\, x + 2\sqrt{w_i}\,(p_i - p_0)\, y + \frac{n_i}{l_i} \Delta(1-\lambda_1) = \sqrt{w_i} \sin^2 \theta_i. \tag{56}$$

Eqs. (56) can be further generalized by introducing a term which accounts for the inaccurate determination of the constant magnitude between successive minima.

Eqs. (56) are solved by successive approximations, assuming some starting k. After that, α_0 is found from (5). Using these k and α_0, we find $p_i^2 - p_0^2$ and $p_i - p_0$, and Eqs. (56) are then solved by the least squares method. Having found x and y, we repeat the solution.

We are thus again faced with Fetlaar's technique, but the elements x and y are found from the condition of minimum sum of squares of deviations of the normal points from the "theoretical" curve, so that in principle the dispersion (rms error) of the elements can be obtained as a byproduct. Calculations using the Kopal—Piotrovski method invariably give unreasonably small dispersions.

The application of the least squares method to Eqs. (56) as long as k and η are not known is not justified, since the deviations of the true k, α_0 from the previous values may be considerable. A more logical course is to apply the least squares method to determine small corrections to the elements, which have been obtained by an alternative method (e.g., Russell's technique). A procedure for the determination of these corrections was developed by Wyse /13/ and Tsesevich /1/.

§29. Determination of corrections to relative elements by the least squares method. Suppose that the approximate values of the relative elements have been found by some simple method.

First let us consider the case of a total eclipse. The tentative elements are used to construct a theoretical light curve, and for each normal point we then calculate the difference between the observed and the theoretical values of l, i.e., $\Delta l_i = l_{obs} - l_{calc}$. Δl can be converted to $\Delta\alpha$ if the depth of the minimum is known. On the other hand, to terms of second order, we have the differential relation

$$\Delta\alpha_i = \frac{\partial\alpha_i}{\partial k}\Delta k + \frac{\partial\alpha_i}{\partial p}\Delta p. \tag{57}$$

The corrections Δk, Δp, like the deviations Δl_i, are generally treated as differentials.

To find Δp, Δk, $\frac{\partial\alpha}{\partial k}$, and $\frac{\partial\alpha}{\partial p}$, we use the standard relations

$$p = \frac{1}{r_2}\sqrt{1-\sin^2 i\cos^2\theta} - \frac{r_1}{r_2},$$
$$\theta = \frac{2\pi(t-t_0)}{P} = \mu(t-t_0),$$
$$k = \frac{r_2}{r_1}.$$

Then

$$\left.\begin{gathered}
\Delta k = \frac{1}{r_1}\Delta r_2 - \frac{k}{r_1}\Delta r_1,\\
\Delta p = \frac{\partial p}{\partial r_1}\Delta r_1 + \frac{\partial p}{\partial r_1}\Delta r_2 + \frac{\partial p}{\partial i}\Delta i + \frac{\partial p}{\partial\theta}\frac{\partial\theta}{\partial t_0}\Delta t_0,\\
\frac{\partial p}{\partial r_1} = -\frac{1}{r_1},\quad \frac{\partial p}{\partial r_2} = -\frac{p}{r_2},\quad \frac{\partial p}{\partial i} = -\frac{\operatorname{ctg} i}{r_2}\left[\frac{1}{r_1+r_2p} - (r_1+r_2p)\right],\\
\frac{\partial p}{\partial\theta} = \frac{\sin^2 i\sin\theta\cos\theta}{r_2(r_1+r_2p)},\quad \frac{\partial\theta}{\partial t_0} = -\frac{2\pi}{P} = -\mu.
\end{gathered}\right\} \tag{58}$$

Inserting from (58) in (57), we get

$$\Delta\alpha_i = A_i\Delta r_1 + B_i\Delta r_2 + C_i\Delta i + D_i\Delta t_0, \tag{59}$$

where

$$\left.\begin{aligned} A &= \left[-\frac{k}{r_1}\frac{\partial\alpha}{\partial k} - \frac{1}{r_2}\frac{\partial\alpha}{\partial p}\right], \quad B = \left[\frac{1}{r_1}\frac{\partial\alpha}{\partial k} - \frac{p}{r_2}\frac{\partial\alpha}{\partial p}\right], \\ C &= -\frac{\operatorname{ctg} i}{r_2}\left[\frac{1}{r_1+r_2p} - (r_1+r_2p)\right]\frac{\partial\alpha}{\partial p}, \quad D = -\frac{\mu\sin^2 i}{2r_2}\frac{\sin 2\theta}{r_1+r_2p}\frac{\partial\alpha}{\partial p}. \end{aligned}\right\} \tag{60}$$

The derivatives $\frac{\partial\alpha}{\partial k}$ and $\frac{\partial\alpha}{\partial p}$ for given k and p can be found by numerical differentiation of the table which lists p as a function of k and α. All the coefficients of the conditional equations (59) can thus be found from the tentative elements. These equations can be generalized to a certain extent by introducing a correction for the depth of the primary minimum. Indeed, since $l_i = (1-\lambda_i)\alpha$, we have

$$\Delta l_i = \alpha_i\Delta(1-\lambda_i) + (1-\lambda_i)\,\Delta\alpha_i$$

or

$$\frac{\Delta l_i}{1-\lambda_i} = \alpha_i\frac{\Delta(1-\lambda_i)}{1-\lambda_i} + \Delta\alpha_i. \tag{61}$$

Eqs. (59) thus take the form

$$\frac{\Delta l_i}{1-\lambda_i} = \alpha_i\frac{\Delta(1-\lambda_i)}{1-\lambda_i} + A_i\Delta r_1 + B_i\Delta r_2 + C_i\Delta i + D_i\Delta t_0. \tag{62}$$

The unknowns in these equations are the corrections $\frac{\Delta(1-\lambda_i)}{1-\lambda_i}$, Δr_1, Δr_2, Δi and Δt_0. Before actually solving these equations, we should determine their weights. The weights depend on the dispersion Δl_i about the theoretical or, equivalently, the average (normal) light curve. It would be improper to assign to these equations some "theoretical" weights, since the accuracy of the different sections of the light curve depends on a variety of random factors, which can hardly be foreseen. A better policy is to find the order of magnitude of the mean errors using a curve plotted from individual observations, and not the normal curve. The individual Δl (for individual observations) fall, with few exceptions, inside the so-called dispersion strip. The shape of this strip enables us to determine the weights of the different sections and of the normal curve. The weight of the equation w_i can be taken inversely proportional to the square of the width of the dispersion strip or, equivalently, to the square of the mean deviation of l from the normal curve. The weights of Eqs. (62) should be calculated only to one significant digit. Determination of weights from the width of the dispersion strip, rather than from the mean square deviations, is preferable, since it is valid also when the observations are nonuniformly distributed along the light curve: after all, the dispersion strip is traced for the entire curve as one whole. To avoid arbitrary decisions in tracing the dispersion strip, it is generally stipulated to contain 90% of all the individual points.

The case of a partial eclipse was also treated by Tsesevich. Suppose that the maximum intensity, as well as the depth of the minima, were

determined with insufficient accuracy. Then for the primary minimum, assuming "large in front," we get

$$\frac{l_1}{l_0} = 1 - \alpha_i L_2, \tag{63}$$

where the exact value of l_0 is unity. Differentiation of (63) yields

$$\alpha \Delta L_2 + L_2 \Delta\alpha = \Delta l_0 - \Delta l_i. \tag{64}$$

Again using Eqs. (58) for $\Delta\alpha$, Δk, and Δp, we get

$$-\Delta l_i = -\Delta l_0 + \alpha_i \Delta L_2 + L_2 \left[-\frac{k}{r_1}\frac{\partial\alpha}{\partial k} - \frac{1}{r_2}\frac{\partial\alpha}{\partial p} \right] \Delta r_1 +$$
$$+ L_2 \left[\frac{1}{r_1}\frac{\partial\alpha}{\partial k} - \frac{p}{r_2}\frac{\partial\alpha}{\partial p} \right] \Delta r_2 - L_2 \frac{\operatorname{ctg} i}{r_2} \left[\frac{1}{r_1 + r_1 p} - (r_1 + r_2 p) \right] \frac{\partial\alpha}{\partial p} \Delta i -$$
$$- L_2 \frac{\sin^2 i \sin 2\theta}{2r_2 (r_1 + r_2 p)} \frac{2\pi}{P} \frac{\partial\alpha}{\partial p} \Delta t_0. \tag{65}$$

Under the U hypothesis, the photometric phase of the eclipse in the secondary minimum is given by

$$\alpha = \frac{1-l}{k^2 L_1} = \frac{1-l}{k^2 (1-L_2)}$$

and

$$\frac{l}{l_0} = 1 - \alpha k^2 (1 - L_2). \tag{66}$$

Differentiation of (66) gives

$$\Delta l_i - \Delta l_0 = \alpha_i k^2 \Delta L_2 - 2\alpha_i (1 - L_2) k \Delta k - k^2 (1 - L_2) \Delta\alpha.$$

Finally, as in the former case, we get

$$-\Delta l_i = -\Delta l_0 - \alpha_i k^2 \Delta L_2 + \left[-\frac{2\alpha}{r_1} - \frac{k}{r_1}\frac{\partial\alpha}{\partial k} - \frac{1}{r_2}\frac{\partial\alpha}{\partial p} \right] k^2 (1 - L_2) \Delta r_1 +$$
$$+ \left[\frac{2\alpha}{r_2} + \frac{1}{r_1}\frac{\partial\alpha}{\partial k} - \frac{p}{r_2}\frac{\partial\alpha}{\partial p} \right] k^2 (1 - L_2) \Delta r_2 -$$
$$- k^2 (1 - L_2) \frac{\operatorname{ctg} i}{r_2} \left[\frac{1}{r_1 + r_2 p} - (r_1 + r_2 p) \right] \frac{\partial\alpha}{\partial p} \Delta i - k^2 (1 - L_2) \frac{\sin^2 i \sin 2\theta}{2r_2 (r_1 + r_2 p)} \frac{2\pi}{p} \frac{\partial\alpha}{\partial p} \Delta t_0. \tag{67}$$

Eqs. (65) and (67) should be solved simultaneously.

The application of the least squares method to find the corrections Δ to tentative elements does not require special skill in the examination of the light curves. This attractive feature, however, is greatly dangerous as it may lead to excessive mathematical idealization of the problem. An inexperienced worker may start with a poor system of elements, which do not fit the light curve and are very far from the true elements, and by a sequence of mathematical manipulations calculate improper corrections. Thus, application of the least squares method to the conditional equations (62), (65), and (67) is appropriate only when the corrections are small, so that terms of second order in Δr_1, Δr_2, etc., can indeed be ignored.

The second danger is the use of "exceptional" points, i.e., those with large Δl_i. The omission of these points is legitimate only if there are pretty good reasons for this or if they exceed by more than a factor of three the mean square deviation (dispersion) from the normal curve and are thus

little probable. On the other hand, it is these "exceptional" points that greatly influence the sum of the squares of deviations from the final theoretical curve and thus the values of the relative elements. In the final analysis, the elements to a certain extent reflect the individual, and fairly arbitrary, decisions of the person doing the calculations, who ignores the "exceptional" points or assigns them relatively low weights. The method is thus highly subjective. These difficulties are not entirely confined to the least squares method in its application to the elements of eclipsing binaries. They are often encountered in the reduction of observational data in various fields.

The solution of the light curves of eclipsing binaries also has a number of specific features which make the application of the least squares method outright invalid, and not only insufficiently rigorous in the mathematical sense. The point is that the deviations Δl_i from the final theoretical curve cannot be treated as random errors. They invariably contain some biased errors and deviations. Even if the light curve is free from biased errors, systematic deviations are inevitable since our working model — a close pair subject to U and D hypotheses — is only a fairly crude approximation to the actual state of things. Δl_i therefore contain a biased component — the so-called model errors /12/. Moreover, even the so-called random part of Δl_i mostly does not follow the law of random errors, since observations carried out during a single night or during one season are not entirely independent and their results clearly depend on the particular constants and the method used in reducing the measurements to some fixed average conditions. The distribution of the random deviations is often represented as a sum of several Gaussians, but this representation is mostly uncertain since the number of observations is limited. The requirement of a minimum sum of squares of deviations is therefore quite arbitrary and the ensuring corrections are most questionable. Fortunately, these corrections generally do not exceed three times the dispersion (so that they can be accepted with a clear conscience) and mostly they are almost equal to the respective dispersions /1/.

The dispersion of the elements is also of considerable interest. Since the random deviations are always larger in amplitude than the biased deviations (otherwise no satisfactory solution can be obtained), the dispersions of the elements are largely determined by the random errors. Since the biased errors are ignored, the dispersions should clearly be considered as the least rms errors of the elements, although this conclusion cannot be rigorously justified. There is unfortunately no exact method for the evaluation of the errors in the elements.

What are the best elements, i.e., elements providing the most faithful representation of the particular system? Suppose that we are dealing with a system whose physical and geometrical characteristics are accurately known. Further suppose that we have idealized the system, representing the stars by ideal spheres with brightness distribution over the disk expressed by (33). Then, by applying mathematical approximation techniques, we should find the radii of these spheres and the x in Eq. (33) which ensure the best fit with the equilibrium figures of stars and with the true brightness distribution over the disk. These ideal elements are called the model elements. Our problem thus reduces to a determination of the model elements with maximum attainable precision. Although no conclusive

solution has been obtained, to first approximation the problem is reduced to a variational problem for the simultaneous determination of two minima:

1) the minimum of the sum of the squares of deviations of the normal points from the theoretical curve, and

2) the minimum of the residual biased deviations, measured by their amplitudes relative to the theoretical curve.

As long as the analytical solution of the problem has not been found, we are reduced to the application of existing methods based on successive approximations (Russell's technique and others). The least squares method should be applied to improve the first tentative elements only if the light curve is highly uncertain and there is no hope to obtain a satisfactory set of best elements. Even in this case, however, the value of the least squares method is mainly as a tool for the determination of the dispersions.

§30. A direct method for estimating the dispersions of relative elements. The dispersions of the relative elements can be found without solving the conditional equations for Δl_i by the least squares method. They can be obtained by a direct technique, differentiating the fundamental equations from which k and α_0 are determined /12/. In practice, however, we need not calculate the dispersion of the elements for every single light curve. As the calculations anyhow give only the order of magnitude of the dispersions, it suffices to estimate them for a number of typical cases of total, partial, and annular eclipses so as to establish the first significant digit which is substantiallly uncertain. Reduction of a large sample of photoelectric light curves by the least squares method (corresponding to a mixture of total, annular, and partial eclipses, the latter with primary minima at least 0.5 deep) showed that r_1 and r_2 are calculated to a few units in the third significant digit, i to a few tenths of a degree, and L_1 and L_2 to a few units in the third significant digit for total eclipses and to within ± 0.01 in partial and annular eclipses. The accuracy of the results obtained from so-called photographic curves (curves from photographic observations) is generally a factor of 2—3 less. If the minima in partial eclipse are shallow ($1 - \lambda_1 < 0.5$), the accuracy markedly deteriorates, and for $1 - \lambda_1 \leqslant 0.1$ the solution is uncertain and it is impossible to decide which of the stars is eclipsed in the primary minimum.

The most objective data on the dispersions of elements are obtained by reducing a number of light curves obtained simultaneously by different observers. In this case, the deviations in the best elements derived from each light curve are truly random, as the data obtained with different instruments are independent. Three light curves are enough for estimating the order of magnitude of the dispersions and the reliability of the average elements.

The dispersions of elements are required mainly in order to appraise the significance of the variations in elements with time, which emerge when light curves corresponding to different epochs are compared. The inaccuracy of the elements should also be taken into consideration when looking for correlations between various elements and physical characteristics of stars (mass, absolute magnitude, spectrum, etc.).

§31. Elliptical orbits (the model of spherical stars). The elements of an elliptical orbit — the eccentricity e and the longitude of the periastron ω — can be found without difficulty only when the eccentricity is small, so that

terms of second and higher order in e can be dropped. The method for the determination of e and ω jointly with the relative elements of the system is also due to Russell.

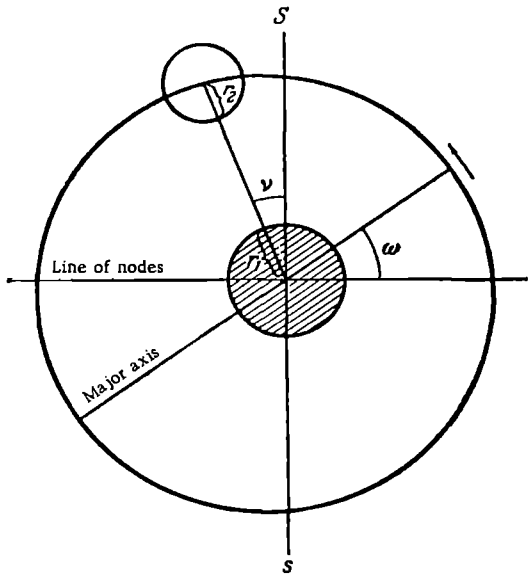


FIGURE 32

Figure 32 shows the orientation of the orbit relative to an earthbound observer for $i = 90°$. The observer's line of sight is along sS. For $i < 90°$ this line is replaced by its projection onto the orbital plane. It is readily seen /1/ that for i fairly close to 90°, the photometric minima occur almost precisely at the conjunctions, i.e., at the time when the apparent distance between the centers of the component stars is minimum. If the angle between the radius-vector r of the orbit and the line sS is reckoned from this line in the direction of motion, the equation of the ellipse has the form

$$r = \frac{1-e^2}{1-e\sin(\nu-\omega)}. \tag{68}$$

The angle ν is the true longitude. The true anomaly is $v = 90° + \nu - \omega$. The true anomaly v and the mean anomaly M are related by the equation of center, which to terms of second order in e is written in the form

$$v = M + 2e\sin M \tag{69}$$

or

$$M = v - 2e\sin v. \tag{70}$$

The corresponding expressions including terms of second order will be found in /1/. For $i = 90°$, Eq. (70) in our case is exact to terms of third order

in e. If M is reckoned from the upper conjunction, i.e., from the eclipse of the brighter component, and the phases of the period are expressed in units of time t, we have

$$M = \mu t + c = \frac{2\pi t}{P} + c. \tag{71}$$

For $v = 0$, $M \neq 0$ and the M corresponding to the phase of the secondary minimum t_2 is

$$M = \frac{2\pi t_2}{P} = \pi - 2e\,[\cos(v - \omega) - \cos\omega], \tag{72}$$

whence

$$\frac{2\pi t_2}{P} = \pi + 4e\cos\omega. \tag{73}$$

Let $e\cos\omega = h$ and $e\sin\omega = g$. Then

$$t_2 - \frac{P}{2} = 2h\frac{P}{2\pi},$$

$$h = \frac{\pi\left(t_2 - \frac{P}{2}\right)}{2P}. \tag{74}$$

Thus h is determined by the deviation of the phase t_2 from the half-period point.

g is much more difficult to find, and yet without this parameter e and ω cannot be determined separately. Following Russell, we write (68) in the form

$$r = 1 - g\cos M + h\sin M. \tag{75}$$

For M at the primary minimum we have

$$M = -2h. \tag{76}$$

Since

$$\theta = \frac{2\pi t}{P},$$

we have

$$\theta = M + 2h. \tag{77}$$

Eqs. (76) and (77) are applicable only for $i \geqslant 80°$. For smaller i, more exact expressions should be used, which contain $\cos^2 i$. Since i is not known beforehand, the solution should anyhow start with Eqs. (74), (76), and (77).

The distance between the centers of the disks (δ) in this case is expressed by

$$\delta^2 = r^2(1 - \cos^2 v\sin^2 i), \tag{78}$$

$$\delta^2 = (1 - g\cos M + h\sin M)^2(1 - \cos^2 v\sin^2 i). \tag{79}$$

Eq. (79) is approximately written in the form

$$\delta^2 = (1 - 2g\cos\theta)\cos^2 i + (1 + 2g\cos\theta)\sin^2 i\sin^2\theta. \tag{80}$$

Eq. (80) can be written in the previously used form

$$\delta^2 = R_1^2(\cos^2 j + \sin^2 j \sin^2\theta), \tag{81}$$

$$R_1 = 1 + g(1 - 2\cos^2 i)\cos\theta, \tag{82}$$

$$\left.\begin{aligned} R_1 \sin j &= (1 + g\cos\theta)\sin i, \\ R_1 \cos j &= (1 - g\cos\theta)\cos i. \end{aligned}\right\} \tag{83}$$

By (83)

$$\operatorname{ctg} j = (1 - 2g\cos\theta)\operatorname{ctg} i. \tag{84}$$

For the secondary minimum, these relations have a slightly different form:

$$\left.\begin{aligned} \nu &= \nu_2, \\ M_2 &= \pi + 2h, \\ \theta' &= M + \pi - 2h = \tfrac{2\pi}{P}(t - t_2) \end{aligned}\right\} \tag{85}$$

(θ' is equivalent to θ reckoned from the phase of the secondary minimum t_2),

$$\delta^2 = R_2^2(\cos^2\beta + \sin^2\beta \sin^2\theta'), \tag{86}$$

$$\left.\begin{aligned} R_2 &= 1 - g(1 - 2\cos^2 i)\cos\theta', \\ \operatorname{ctg}\beta &= (1 + 2g\cos\theta')\operatorname{ctg} i. \end{aligned}\right\} \tag{87}$$

We see from these expressions that if R_1 and R_2 are approximately taken as constant, the eclipse in the primary minimum is the result of motion in a circle of radius R_1 and inclination j, and the eclipse in the secondary minimum is the result of motion in a circle of radius R_2 and inclination β.

Since the variation of $\cos\theta$ at the two minima is small compared to the variation of $\sin\theta$, it can be replaced by its average value η:

$$\eta = \int_{-\zeta}^{+\zeta} \cos\theta\, d\theta : \int_{-\zeta}^{+\zeta} d\theta = \frac{\sin\zeta}{\zeta} \approx \frac{1}{3}(2 + \cos\zeta), \tag{88}$$

where ζ is the value of θ (or θ') at the beginning or at the end of the corresponding minimum.

Since the radii of the components are determined assuming an orbit of unit radius (in our case, R_1 and R_2), the two minima clearly give different radii for the two components (the corresponding values are primed):

$$\left.\begin{aligned} r_1' &= \tfrac{r_1}{R_1}, \quad & r_2' &= \tfrac{r_2}{R_1}, \\ r_1'' &= \tfrac{r_1}{R_2}, \quad & r_2'' &= \tfrac{r_2}{R_2}. \end{aligned}\right\} \tag{89}$$

Here r_1 and r_2 are expressed in units of the semimajor axis of the ellipse. k is hardly affected by the ellipticity of the orbit.

Finally we get:

for the primary minimum

$$\left.\begin{aligned} r_1' &= r_1(1 - g\eta), \\ \operatorname{ctg} j &= \operatorname{ctg} i\,(1 - 2g\eta); \end{aligned}\right\} \tag{90}$$

for the secondary minimum

$$\left.\begin{aligned} r_1'' &= r_1(1+g\eta), \\ \operatorname{ctg}\beta &= \operatorname{ctg} i\,(1+2g\eta). \end{aligned}\right\} \tag{91}$$

If it were possible to find r_1 and i for the two minima separately, we could calculate the product $g\eta$:

$$g\eta = \frac{r_1''-r_1'}{r_1''+r_1'}, \qquad g\eta = \frac{1}{2}\,\frac{\operatorname{ctg} i''-\operatorname{ctg} i'}{\operatorname{ctg}\beta+\operatorname{ctg} i}. \tag{92}$$

Both these equalities should give the same value of $g\eta$ within the accuracy of our calculations and our solution (the accuracy of the curve and the validity of the model). Before Eqs. (92) can be used to calculate $g\eta$, we require r_1 and i separately for each minimum. In practice, these parameters can be found directly only when the eclipse in one of the minima is total, both minima are almost of the same depth, and the U hypothesis provides an adequate approximation. If the depth of the secondary minimum is small, the inaccuracy in the determination of $g\eta$ due to a number of side factors (e.g., differences of limb darkening for the two stars) will be so large that the resulting $g\eta$ can hardly be considered significant.

In case of a partial eclipse, for $g \neq 0$, α_0 (or α_0') is no longer equal for the two minima. If the U hypothesis is applicable, Eq. (5) takes the form

$$\frac{1-\lambda_1}{\alpha_{1.0}} + \frac{1-\lambda_2}{k^2\alpha_{2.0}} = 1. \tag{93}$$

Russell suggested solving Eq. (93) graphically or by successive approximations simultaneously with the two equations which are obtained when the numerical values of the functions $\chi(k,\ \alpha_{1.0};\ 0.25)$ and $\chi(k,\ \alpha_{2.0};\ 0.25)$ are determined from the shape of the two minima. We are not particularly fond of this technique, since the χ for the secondary minimum is found with a high uncertainty. To solve Eq. (93), $\alpha_{1.0}$ and $\alpha_{2.0}$ should be linked up through g. Clearly, $\alpha_{1.0}$ and $\alpha_{2.0}$ differ only because the $\varrho_{1.0}$ (replacing ϱ_0) of the primary minimum is different from $\varrho_{2.0}$. Let the difference in ϱ be $2\Delta\varrho$:

$$\Delta\varrho = \frac{\Delta\delta}{r_1} \approx \frac{g\cos i}{r_1}. \tag{94}$$

For $\alpha_{1.0}$ and $\alpha_{2.0}$ we then have

$$\left.\begin{aligned} \alpha_{1.0} &= \alpha_0 + \left(\frac{\partial\alpha}{\partial\varrho}\right)_{\alpha=\alpha_0}\Delta\varrho, \\ \alpha_{2.0} &= \alpha_0 - \left(\frac{\partial\alpha}{\partial\varrho}\right)_{\alpha=\alpha_0}\Delta\varrho, \end{aligned}\right\} \tag{95}$$

or

$$\left.\begin{aligned} \alpha_{1.0} &= \alpha_0 + \left(\frac{\partial\alpha}{\partial\varrho}\right)_{\alpha=\alpha_0}\frac{g\cos i}{r_2}, \\ \alpha_{2.0} &= \alpha_0 - \left(\frac{\partial\alpha}{\partial\varrho}\right)_{\alpha=\alpha_0}\frac{g\cos i}{r_2}. \end{aligned}\right\} \tag{96}$$

Since the elements are known to first approximation, Eqs. (96) are readily applied to calculate the corrections to some average value α_0 corresponding to $g = 0$ or to reduce $\alpha_{2.0}$ to $\alpha_{1.0}$:

$$\alpha_{2.0} = \alpha_{1.0} - 2\left(\frac{\partial\alpha}{\partial\varrho}\right)_{\alpha=\alpha_0}\frac{g\cos i}{r_2}. \tag{97}$$

Combining (93) with $\chi(k, \alpha_{1.0}; 0.25)$ we can then find k and α_0. After that we can proceed with the second approximation; in practice, however, the accuracy of the elliptical elements is fairly low so that the first approximation is quite adequate.

§32. The motion of the line of apsides. The presence of a third body. In some cases the deviation of the secondary minimum from the phase $\frac{P}{2}$ is variable. Therefore $e\cos\omega$ is also variable. These variations follow a regular sine curve, and evidently the variation of ω is the result of the uniform precession of the line of apsides. Inspection of the light variation elements shows that both minima are displaced. The variation of the epochs of the primary minimum is in the opposite direction relative to the variation of the epochs of the secondary minimum.

The precession of the line of apsides is mainly caused by the tidal deformation of the stars. In this case the rate of precession ω' depends on the moments of inertia of the stars relative to the three principal axes of symmetry of their equilibrium figures. These moments of inertia, in their turn, are determined by the distribution of density in the stellar interior. Homogeneous density leads to a large ω', whereas concentration of mass toward the center of the star lowers ω', since the star in this case can be considered as an ideal point mass. Sometimes the precession of the line of apsides and the variation of other orbital elements may be caused by the presence of a third body in the system.

The motion of the line of apsides is most prominent in systems with orbits of moderate eccentricity. There are excellent determinations of ω' in YCyg, RU Mon, GL Car, and fairly reliable determinations for five other binaries.

A direct indication of the presence of a third star near a close binary is provided by a periodic synchronous fluctuation of the epochs of the two minima. This is an outcome of the so-called light equation. The third star is generally relatively far from the close binary, so that to fair approximation the center of inertia of the close binary traces an elliptic orbit about the center of inertia of the entire triple system. Although the dimensions of this orbit are generally small — it is smaller than the Earth's orbit around the Sun — the distance of the binary system from the Earth will nevertheless fluctuate by some tens of millions of kilometers as its center of inertia describes the orbit. The epochs of the minima depend on the position of the center of inertia of the binary in the orbit. It takes a few minutes for the light to traverse the diameter of the orbit. This effect of the finite velocity of light in the epochs of minima is known as the light equation. The variation of the light equation can be obtained by plotting against time the deviations of the epoch of the minimum from the epoch calculated assuming a constant period. The periodic fluctuations can be reduced to a single period; the light equation is considered within that period, since the orbital period of the center of inertia of the close binary remains constant (being equal to the orbital period of the third body). Figure 33 is the light equation of the variable RT Per according to Dugan. For this system the period of the third body is 16,000 P (P is the period of the close binary). The orbital elements of the center of inertia of the close binary are calculated by the same technique as the orbital elements of a spectroscopic binary, since the

light equation characterizes the motion along the observer's line of sight. A useful method for the solution of this problem was proposed by Woltjer /1/.

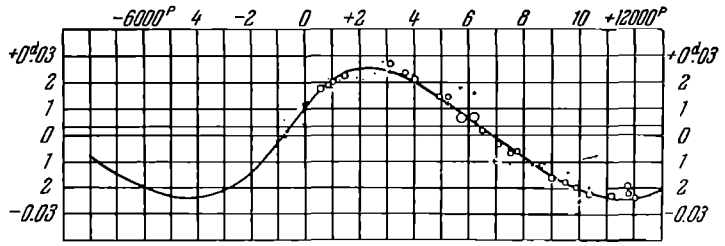


FIGURE 33

For some triple systems, such as β Per, λ Tau, and VV Ori, the orbit of the center of mass of the close binary can be derived from the fluctuations in γ (the radial velocity of the center of mass of the binary).

Since the study of triple systems is a particular problem, and they relatively are few, we will not go into this topic.

§33. **Tidal deformations (the model of ellipsoidal stars).** Tidal forces produce considerable deformation of stars in close binaries. The deformation is the largest when the orbital period of the pair is equal to the period of axial rotation of the star. In this case we are dealing with the so-called classical problem of binary stars, first solved by G. Darwin for homogeneous incompressible fluid bodies. Compressibility reduces the tidal deformations. For a polytropic gas, where the pressure p is a function of density,

$$p = K\varrho^{1+\frac{1}{n}}, \tag{98}$$

K and n being constant, the equilibrium figures of the close binary components were calculated by Chandrasekhar. Since the gas in normal stars (the white dwarfs excluded) behaves as an ideal gas

$$p = \frac{R}{\mu}\varrho T$$

(T is the absolute temperature, R is the gas constant, μ is the mean atomic weight), Eq. (98) reduces to the requirement that T is only a function of ϱ, specifically

$$T \sim \varrho^{\frac{1}{n}}. \tag{99}$$

In the model of a classical binary, p is only a function of ϱ /13/, and (99) can therefore be used as an interpolation formula for calculating the density distribution in the stellar interior. Eq. (98) virtually covers all the possible cases of density distribution. The case $n = 0$ corresponds to homogeneous incompressible stars, and $n = 5$ corrresponds to Roche's model when almost the entire mass is concentrated at the center of the star. $n = 0$ and $n = 5$ are the two extreme cases. The true density distribution in a star, in accordance with our current notions of the constitution of stellar interiors, is expressed to fair approximation by a polytrope with $n = 3$ or

$n = 4$. In either case the external surface of the star (its equilibrium figure) is very close to an equipotential surface (for the combined potential of centrifugal and gravitational forces) of Roche's model. Therefore, Roche's model can be successfully applied in theoretical estimates of the tidal deformation of real stars.

The equilibrium figure for each of the component stars is slightly pear shaped. It can be approximately represented by a triaxial ellipsoid. Let ε be the eccentricity of the equatorial section of this ellipsoid (the section in the orbital plane) and ε' the eccentricity of the principal meridional section (the section through the spin axis and the line of centers of the component stars). ε and ε' can be expressed by

$$\varepsilon_1^2 = 3q_1 \frac{m_2}{m_1} v_1^3, \tag{100}$$

$$(\varepsilon_1')^2 = q_1 \left(1 + 4\frac{m_2}{m_1}\right) v_1^3. \tag{101}$$

Here the subscript 1 qualifies the star being considered, the subscript 2 identifies the second, perturbed component of the binary. v_1 is the mean radius of the star in the absence of centrifugal and tidal forces. Approximately it can be taken equal to $(abc)^{1/3}$, where a, b, c are the ellipsoid semiaxes. The factor q depends on the internal constitution of the star. For real stars q is close to 1, and for the homogeneous incompressible model $q = 2.5$.

The ellipsoidal figure of the components destroys the constancy of the magnitude between successive eclipses. The maximum, as we see from Figure 28, is attained for $\theta = 90°$ and $\theta = 270°$. This also follows from the variation of the brightness of a rotating ellipsoid. The area of the projection of the ellipsoid on the plane of the sky is given by /1/

$$S = \pi a^2 \cos i \left\{\left(\frac{b}{a}\right)^2 + \left(\frac{c}{a}\right)^2 \operatorname{tg}^2 i \left[\left(\frac{b}{a}\right)^2 \cos^2\theta + \sin^2\theta\right]\right\}^{\frac{1}{2}}, \tag{102}$$

where a, b, c, i, θ are defined as before. We see from (102) that S, and hence the brightness of the ellipsoid, depend on $\left(\frac{c}{a}\right)^2 \tan^2 i$, and not on $\left(\frac{c}{a}\right)$ alone.

Let $\frac{c}{a} = u\frac{b}{a}$ $(u < 1)$. To simplify (102), we write

$$\frac{c}{a}\operatorname{tg} i = \frac{b}{a}\operatorname{tg} i', \quad \operatorname{tg} i = \frac{b}{c}\operatorname{tg} i', \tag{103}$$

replacing the true inclination i by some hypothetical inclination i'. From (102) we thus get

$$S \sim (1 - \varepsilon^2 \sin^2 i' \cos^2\theta)^{\frac{1}{2}}. \tag{104}$$

This is the expected dependence of S on ε for a star in the form of an oblate spheroid with semiaxes a and b, rotating around the b axis. For the product $\varepsilon^2 \sin^2 i'$ we write z. The maximum is observed when $\cos\theta = 0$. Let the maximum brightness of each star be L_1 and L_2, so that $L_1 + L_2 = 1$. The brightness of two synchronously rotating ellipsoids is then given by

$$l = L_1(1 - z_1\cos^2\theta)^{\frac{1}{2}} + L_2(1 - z_2\cos^2\theta)^{\frac{1}{2}}. \tag{105}$$

For z_1 fairly close to z_2, we may take $z_1 = z_2 = z$. Eq. (105) then takes the form

$$l = (1 - z\cos^2\theta)^{\frac{1}{2}}. \tag{106}$$

Consider the U hypothesis. When the smaller of the two stars is eclipsed by the large star, we get

$$l = (1 - z\cos^2\theta)^{\frac{1}{2}}[L_1 + (1-\alpha)L_2]. \tag{107}$$

When the large star is eclipsed by its smaller companion, (107) is replaced by

$$l = (1 - z\cos^2\theta)^{\frac{1}{2}}[L_1(1 - k^2\alpha) + L_2]. \tag{108}$$

We see that the brightness is always proportional to the factor $f = (1 - z\cos^2\theta)^{1/2}$. Dividing all l through by f, we obtain a rectified light curve, which is no longer dependent on f. Between the eclipses, the rectified curve has a constant phase, as in the case of the spherical star model. Since the ellipsoids are similar, their projections on the plane of the sky are similar to ellipses with the semiaxes b_1, $d_1 = a_1 f$ and b_2, $d_2 = a_2 f$ (Figure 34). Since both ellipses can be considered as projections of two intersecting circles with radii d_1 and d_2 onto some inclined (to the plane of the sky) plane, the α for the ellipses is the same as for the circles. The light variation in the minima of the rectified curve is the same as for spherical stars of variable radii, proportional to f. z can be found without difficulty. Suppose that although the brightness for $\theta = 90°$ and $\theta = 270°$ is unity by definition, we made an error of a in its determination. The true brightness is then $\frac{l'}{1+a} = l$ (l' is the brightness of the normal points from the light curve). Squaring (106), we get

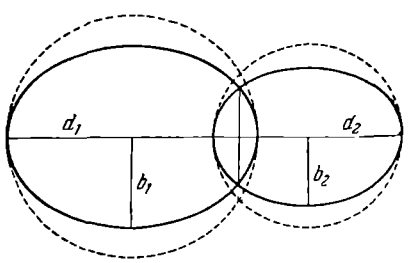


FIGURE 34

$$\frac{l'^2}{1+2a} = 1 - z\cos^2\theta, \qquad z(1+2a)\cos^2\theta - 2a = 1 - l'^2. \tag{109}$$

For each normal point we thus obtain a linear equation with the unknowns $z(1 + 2a)$ and $2a$. Solving these equations by the least squares method, we readily find the value of z and the correction a. First it is advisable to plot (109) in the coordinates of $1 - l'^2$ vs. $\cos\theta$ and establish that there are departures from the linear variation of $1 - l'^2$ as a function of $\cos^2\theta$. The deviations are observed for normal points affected by the eclipse. These points should be omitted from the sample used in calculating z.

When z has been found, the elements of the system can be determined. A convenient method of solution was first proposed by Russell /2/. For δ we write

$$\delta^2 = \cos^2 i + \sin^2 i \sin^2\theta = d_1^2\varrho^2 = a_1^2 f^2 \varrho^2. \tag{110}$$

For a total eclipse we have as before

$$\left.\begin{aligned} A+B\sin^2\theta_n &= f_n^2\varrho_n^2,\\ A+B\sin^2\theta_{0.6} &= f_{0.6}^2\varrho_{0.6}^2,\\ A+B\sin^2\theta_{0.9} &= f_{0.9}^2\varrho_{0.9}^2. \end{aligned}\right\} \tag{111}$$

Eliminating A and B between these equations, we get

$$\frac{\varrho_n^2-\varrho_{0.6}^2}{\varrho_{0.6}^2-\varrho_{0.9}^2}=\frac{(\sin^2\theta_n-\sin^2\theta_{0.6})\,f_{0.9}^2}{(\sin^2\theta_{0.6}-\sin^2\theta_{0.9})\,f_n^2}. \tag{112}$$

This relation replaces Eq. (17). Further calculations proceed along the same lines as for spherical stars. A and B are obtained from

$$\left.\begin{aligned} A &= \frac{\cos^2 i}{a_1^2}, \qquad B=\frac{\sin^2 i}{a_1^2},\\ f_0^2(1+k)^2 &= A+B\sin^2\theta_0,\\ f_{1.0}^2(1-k)^2 &= A+B\sin^2\theta_{1.0} \end{aligned}\right\} \tag{113}$$

$\sin^2\theta_0$ and $\sin^2\theta_{1.0}$ are obtained as before once k has been found:

$$\left.\begin{aligned} \sin^2\theta_0 &= \frac{\sin^2\theta_{0.6}f_{0.9}^2+(\sin^2\theta_{0.6}-\sin^2\theta_{0.9})(1-z)\psi_0(k)}{f_{0.9}^2-(\sin^2\theta_{0.6}-\sin^2\theta_{0.9})z\psi_0(k)},\\ \sin^2\theta_{1.0} &= \frac{\sin^2\theta_{0.6}f_{0.9}^2+(\sin^2\theta_{0.6}-\sin^2\theta_{0.9})(1-z)\psi_{1.0}(k)}{f_{0.9}^2-(\sin^2\theta_{0.6}-\sin^2\theta_{0.9})z\psi_{1.0}(k)}. \end{aligned}\right\} \tag{114}$$

The solution for the case of a partial eclipse also proceeds along the same lines as before. Forming the differences

$$\frac{\varrho_n^2-\varrho_0^2}{\varrho_m^2-\varrho_{0.0}^2}=\frac{f_m^2\sin^2\theta_n}{f_n^2\sin^2\theta_m}, \tag{115}$$

we find that the relevant expressions are again the same as for spherical stars, provided that $\sin\theta_n$ and $\sin\theta_m$ are replaced by the reduced functions

$$\sin\theta_n'=\frac{\sin\theta_n}{f_n}, \qquad \sin\theta_m'=\frac{\sin\theta_m}{f_m}. \tag{116}$$

After that the calculations follow Russell's technique or any of the alternative methods. Once k and α have been found, A and B are determined from the relations

$$\left.\begin{aligned} f_0^2(1+k)^2 &= A+B\sin^2\theta_0,\\ f_{1.0}^2[1+kp(k_1\alpha_0)]^2 &= A. \end{aligned}\right\} \tag{117}$$

The D solution is no different from the U solution. In this case, however, z cannot be taken equal to the value obtained from the rectified light curve. The expression for the brightness of the system between eclipses is

$$l=1-Z\cos^2\theta, \tag{118}$$

where

$$Z=\tfrac{4}{5}z+\tfrac{16}{175}z^2+\ldots \tag{119}$$

In practice, since z is determined with high uncertainty, we may take

$$z=\tfrac{5}{4}Z. \tag{120}$$

This is the z to be used in further calculations.

§34. The sphere—ellipsoid model. The model of similar ellipsoids is inapplicable to numerous close binaries. The mass ratio of the components is often far from unity, and the mean radius of the less bulky star is much greater than the mean radius of the massive component. Consider a system where $m_2 = 2m_1$, $\nu_2 = 0.2$, $\nu_1 = 0.4$. For the components of this system we have by (100)

$$\varepsilon_2^2 \approx 0.012 q_2, \quad \varepsilon_1^2 = 0.384 q_1.$$

It is evident that ε_2^2 is negligible compared to ε_1^2 and the more bulky star can be treated as a sphere. In some cases (U Cep, and others) the axial period of rotation of the smaller and brighter component is not equal to its orbital period. In this case the spherical approximation is even more valid. In fact, however, the star is not an exact sphere but an oblate spheroid of revolution. The effect of the spheroidal shape on the light curve cannot be taken into consideration at this stage.

In the sphere—ellipsoid model /16/ the smaller and brighter star is regarded as a sphere, and its larger companion is an ellipsoid. For this model of a close binary, Eq. (105) is replaced by the equality

$$l = L_1 (1 - z_1 \cos^2 \theta)^{\frac{1}{2}} + L_2, \tag{121}$$

and (107) changes to

$$l = L_1 (1 - z_1 \cos^2 \theta)^{\frac{1}{2}} + (1 - \alpha) L_2. \tag{122}$$

To rectify the light curve, we should find a tentative solution taking $z_1 = z_2 = 0$, say, or using the model of similar ellipsoids. This solution, generally accurate to two significant digits, gives correct values of L_1 and L_2. Then (121) is rewritten in the form

$$(l - L_2)^2 = L_1^2 (1 - z_1 \cos^2 \theta). \tag{123}$$

Eq. (123), like Eq. (109), is linear in z_1. With $L_1 < 0.1$, the light curve can be rectified if z_1 is accurate to within ± 0.1. In this case rectification entails adding to all l the quantity $L_1 (1 - f)$ (and not dividing l by f as in the previous model). The right-hand side of (122) thus reduces to the form $L_1 + (1 - \alpha) L_2$.

In subsequent calculations the ellipse with semiaxes b_1, d_1 is replaced by the inscribing circle of radius d_1 (Figure 35). Sofronitskii's calculations have shown that this substitution virtually does not affect the α. The entire effect of different equilibrium figures of the two stars thus reduces to a variation of the radius of the large component with phase. The dependence of ϱ on f is the same as in the model of similar ellipsoids, but now k is a function of f:

$$k' = \frac{k}{f}, \quad k = \frac{a_2}{a_1}. \tag{124}$$

This somewhat complicates further solution of the problem. The point is that the functions $\varrho^2 (k', \alpha)$ in Eqs. (112) and (115) correspond to different k'. The reduction of all ϱ^2 to a common k_0', corresponding to $\theta = 0$ for the case of a total eclipse, can be achieved in the following way. First note that

the k' for $\theta = \theta_{0.9}$ and $\theta = \theta_{0.6}$ are virtually equal to k_0' for all z. The deviation Δk of k_0' from k' is significant only for $n = 0.1, 0.2, 0.3$. For these points we should thus find suitable corrections to $\varrho^2(k_0', \alpha)$, which are substituted for $\varrho^2(k', \alpha)$ in (112). Dadaev /17/ calculated the appropriate corrections to $\psi_n(k')$ for various k' and z for the U and D hypotheses in the form

$$\Delta\psi_n(k') = \Phi_n(k')\,\Delta k_n', \tag{125}$$

where

$$\Delta k_n' = \left(\frac{1 - z\cos^2\theta'}{1 - z\cos^2 U_n}\right)^{\frac{1}{2}} - 1. \tag{126}$$

Here for θ' we may take $\theta = \frac{\theta_{0.0} + \theta_{0.6}}{2}$. This is a somewhat better approximation than $\theta = 0$. Subtracting $\Delta\psi_n(k')$ from the numerical values of $\psi_n(k')$, calculated in the usual way from $\sin^2\theta$, we obtain the values of $\psi_n(k_0')$ corresponding to a common k_0', which is given by

$$k_0' = \frac{k}{(1 - z\cos^2\theta')^{\frac{1}{2}}}. \tag{127}$$

Since the true z may greatly deviate from those calculated from (100), the problem should be solved for various values of z (0.1, 0.2, 0.3, 0.4, 0.5). For every given z, we find the corrections $\Delta\psi_n(k')$ and from tables take k_0' and k. After that the best possible fit to the light curve is found and the differences $O - C$ (obs. –calc.) are determined. The z which gives the best fit of the light curve is selected as the true z. The accuracy of the z determined in this way is not particularly high. Even the best light curves give z to within ±0.05. Commonly, the z determined from good (photoelectric) light curves is accurate to within ±0.1.

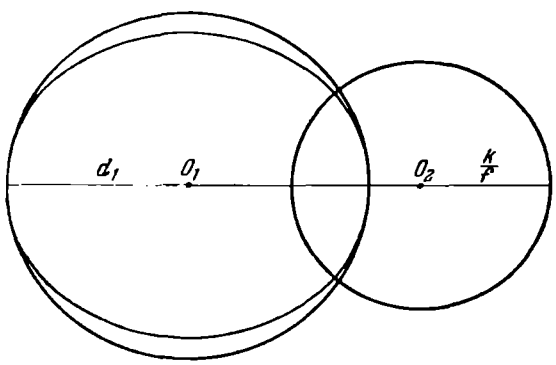


FIGURE 35

In case of a partial eclipse the errors in z rapidly increase with decreasing α_0, and the best policy is therefore to take z from Eq. (100). χ and ψ should be calculated for the same k_0', (127), taking $\theta' = \theta_{0.5}$. It

should be remembered that this procedure only gives an approximate solution. However, if the weaker component has a large z, the sphere—ellipsoid model generally gives a better fit of the light curve than the model of spherical stars or the model of similar ellipsoids with z calculated from the part of the light curve between the eclipses.

A, B, and other elements are obtained as in the model of similar ellipsoids.

§35. **The phase effect.** The phase effect generally represents the influence of the light of one component reflected from the surface of its companion on the light curve of a close binary. We are in fact dealing not with reflection but with absorption followed by re-emission of light quanta. The albedo for integrated light re-emission is unity, since by the second law of thermodynamics the energy flux incident on the surface of a star cannot penetrate into the hot interior. It only produces a local increase of surface temperature. The atmospheres of both stars contribute to re-emission, but this effect is noticeable in the light curve only if one of the stars is much brighter than the other. The variation of brightness with phase is shown in Figure 36. Near the primary minimum, the re-emitting part of the satellite's surface is virtually invisible to the earthbound observer and thus does not increase the brightness of the system. Near the secondary maximum, on the other hand, it is almost completely visible. The increase of magnitude from the primary to the secondary minimum (toward $\theta = \pi$) represents the phase effect. At the maximum $\theta = \pi$ (ignoring the effect of the eclipse for the moment), the corresponding increment is $L_r(\pi)$. For the satellite (subscript 1) we have according to Eddington

$$L_{r,1}(\pi) = \frac{17}{24} L_2 \left(r_1^2 + \frac{3}{4} r_1^3 + \dots \right). \qquad (128)$$

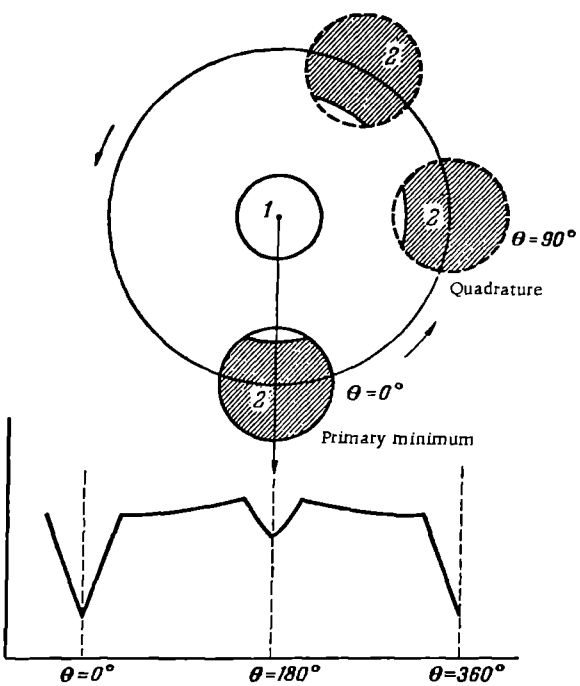


FIGURE 36

For $\theta < \pi$, the increment $L_{r,1}(\theta)$ is smaller. Eddington obtained

$$L_{r,1}(\theta) = L_{r,1}(\pi)\frac{\sin\theta - \theta\cos\theta}{\pi}. \tag{129}$$

Milne derived a more complex expression for this function, which however numerically is close to (129). Pike proposed the expression

$$L_{r,1}(\theta) = L_{r,1}(\pi)\sin^n\frac{\theta}{2} \quad (n = 4.5). \tag{130}$$

Krat derived an empirical formula

$$L_{r,1}(\theta) = L_{r,1}(\pi)\sin^{4.2}\frac{\theta}{2}. \tag{131}$$

The three equations — (129), (130), and (131) — are virtually equivalent. The maximum divergence between the three is observed for $\theta = 90°$ For this θ, the coefficient of $L_{r,1}(\pi)$ in (129) is 0.32, that in (130) is 0.21, and that in (131) is 0.23. Changing the power index in (131) from 4.2 to 4.0, we change this coefficient by a mere 0.02 (it increases to 0.25). Since $L_{r,1}(\pi)$ is anyhow small (it never exceeds 0.1), these differences fall within the margin of error of l. Taking $n = 4$ in (130), we get

$$L_{r,1}(\theta) = L_{r,1}(\pi)\left[\frac{1}{4}(1 - 2\cos\theta + \cos^2\theta)\right]. \tag{132}$$

For the second component, we should clearly replace θ with $\pi - \theta$ in (132). Then

$$L_{r,2}(\theta) = L_{r,2}(0)\left[\frac{1}{4}(1 + 2\cos\theta + \cos^2\theta)\right]. \tag{133}$$

The overall phase effect is given by

$$L_{r,1}(\theta) + L_{r,2}(0) = \frac{L_{r,1}(\pi) + L_{r,2}(0)}{4} + \frac{L_{r,1}(\pi) + L_{r,2}(0)}{4}\cos^2\theta - \frac{1}{2}[L_{r,1}(\pi) - L_{r,2}(0)]\cos\theta. \tag{134}$$

Putting

$$b_1 = \frac{L_{r,1}(\pi) + L_{r,2}(0)}{4}, \quad b = \frac{1}{2}[L_{r,1}(\pi) - L_{r,2}(0)],$$

we get

$$L_{r,1}(\theta) + L_{r,2}(\theta) = b_1 + b_1\cos^2\theta - b\cos\theta. \tag{135}$$

We see from (134) that the observed phase effect is determined by the difference in the re-emission of the two components at $\theta = \pi$ (the primary star) and $\theta = 0$ (the satellite). Moreover, the term $b_1\cos^2\theta$ may somewhat obscure the ellipsoidal effect, since it is maximum for $\theta = 0$, π and vanishes for $\theta = 90°$, $270°$. If $L_{r,1}(\pi) \approx L_{r,2}(0)$, and the ellipsoidal effect is zero, two shallow minima will form at $\theta = 90°$, $270°$.

The phase effect is considered jointly with the ellipsoidal effect, since in this case the system components are close to each other and the tidal deformation is considerable. For small z, the light variation between the eclipses is given by

$$l = a + 1 - \frac{z}{2}\cos^2\theta + b_1 + b_1\cos^2\theta - b\cos\theta$$

or

$$l = a_0 - c\cos^2\theta - b\cos\theta, \tag{136}$$

where

$$a_0 = a + 1 + b_1, \quad c = \frac{z}{2} - b_1,$$

and a is interpreted as the zero point of the light scale, as before. Solving Eqs. (136) by the least squares method, we find a_0, c, and b. According to Russell, the maximum is selected so that $L_1 + L_2 = 1$ as before, but L_1 also incorporates $2b$. The phase effect is not eliminated by subtraction from the light curve but by adding to l the correction

$$\frac{1}{2}[L_{r,1}(\pi) + L_{r,2}(0)] + b\cos\theta.$$

Subtraction is of course a more natural approach, but elevation of the zero point followed by addition of the above binomial also involves reduction of contrast between the primary minimum and l with $\theta = \pi$, as in the subtraction. The two approaches are thus equivalent in this case.

If $L_{r,2}(0)$ is not known, we can ignore this quantity, and l corrected for the phase effect (l_1) takes the form

$$l_1 = \frac{l + b(1+\cos\theta) - b_1\cos^2\theta}{a_0 + b}. \tag{137}$$

These l_1 are rectified for the ellipsoidal effect (similar ellipsoids) by dividing through by f, which gives the final rectified values

$$l_{\text{rect}} = \frac{l_1}{f}. \tag{138}$$

For the sphere—ellipsoid model, we get from (121) to second order in z

$$l = 1 - \frac{z_1}{2} L_1\cos^2\theta,$$

and the ellipsoidal correction is simply added to l_1:

$$l_{\text{rect}} = l_1 + \frac{z_1}{2} L_1\cos^2\theta,$$
$$l_{\text{rect}} = l_1 + (c + b_1)\cos^2\theta. \tag{139}$$

The re-emitted light gives a total limb darkening which does not fit the total limb darkening prescribed by the D hypothesis. The disk brightness in re-emitted light for $\theta = \pi$ is virtually zero at a distance of 0.9 of the radius from the center. Therefore, in case of a partial eclipse in the primary minimum, the part of the satellite disk illuminated by the primary may be completely occulted. This is generally the case when $\alpha_0 > 0.7$. Then, seeing that in the primary minimum the larger star is eclipsed, we have for the U hypothesis

$$1 - \lambda_1 = \alpha_0 L_2,$$
$$1 - \lambda_2 = \alpha_0 k^2 [L_1 - L_{r,1}(\pi)] + L_{r,1}(\pi).$$

Eq. (5) takes the form

$$\frac{1}{1 - L_{r,1}(\pi)}\left[1 - \lambda_1 + \frac{1-\lambda_2}{k^2}\right] = \alpha_0. \tag{140}$$

A similar equation is obtained for the D hypothesis. The $L_{r,1}(\pi)$ for these systems is taken equal to $2b$ with sufficient accuracy. The use of theoretical relations (128) in the calculation of $L_{r,1}(\pi)$ and $L_{r,2}(0)$ is not justified, since the observed differences of these parameters greatly deviate from the theoretical figures. The reason for this is apparently the peculiar structure of stellar atmospheres. At present it is almost impossible to find $L_{r,2}(0)$ from the difference $L_{r,1}(\pi) - L_{r,2}(0)$ using (128). A reliable treatment of the phase effect is possible only when $L_{r,2}(0)$ is known to be small.

§36. Non-stable stars. The early observers of β Lyr noted irregularities in its light curve and various changes in the spectrum (Blagg, Maury, and others). Later, when the accuracy of photometric measurements increased (photoelectric observations of Stebbins, Huffer, and others), it became clear that the light curve of this variable is highly asymmetric in the primary maximum. After that, numerous other stars with similar properties were discovered (U Cep, RV Oph, RS CVn, AO Cas, and others). Later O. Struve established the principal reason for the skewness of the light curve of most eclipsing variables of this type. Streams of gas were observed to be ejected from the atmospheres of one or both components, forming a gaseous nebula around the close binary. This gaseous envelope as a rule expands and dissipates in space (β Lyr). To better understand this effect, consider a family of equipotential surfaces around the stars. These exterior equipotential surfaces are virtually insensitive to the density distribution in the interior of the star. Their configuration can thus be described with fair approximation by Roche's model. The intersection of these surfaces with the orbital plane for a model with component masses in a ratio of 2:3 is shown in Figure 37 according to Struve and Huang Su-Shu /18/. The points L_1, L_2, and L_3 are the critical points of Roche's model (or Lagrange's points).

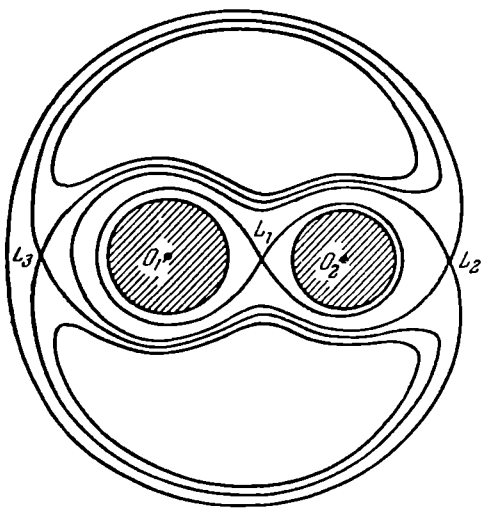


FIGURE 37

A similar configuration is obtained when the equipotential surfaces are cut by any other plane through the line of centers O_1O_2. The gas may move freely along the equipotential surfaces, without doing any work against the force of gravity. If for some reason the stars expand and reach their Roche surface limits, which are joined at the point L_1, further expansion will lead to the formation of some common atmosphere bounded by the equipotential surface through the point L_2. If only one of the stars expands, a jet of gas may flow from its atmosphere to the atmosphere of the quiescent component in the direction of the gravity vector. The second component will not remain quiescent for long /19/, since the increase in its mass will increase the pressure and temperature at the center, greatly accelerate the rate of thermonuclear reactions,

raise the energy flux radiated by the star (its luminosity), and inevitably lead to its expansion. This is, however, an entirely different problem. The expansion of the star is an accelerated process. The rate of expansion increases with time. As soon as the ejection velocity exceeds the critical (parabolic) value at the star's surface, the gas jet will move across the equipotential surfaces away from the two stars. In this case, the two stars will not be able to build up a common atmosphere. Nevertheless, as we are dealing with gas flow, and not with individual particles, the stream will always diverge over the equipotential surfaces, thus deflecting toward the quiet component. Since the stream can flow unimpeded toward the quiet component only after the expanding gas has reached the point L_1, where the specific angular momentum (the angular momentum per unit mass) is very small (this point being close to the center of inertia), the stream beyond that point will necessarily be deflected from the line of centers in the direction opposite to the sense of rotation (conservation of the angular momentum). Henceforth the expansion will be almost radial relative to the center of inertia of the system. At some distance from the line of centers, the potential has a second maximum (the first maximum is situated on the star's figure), on the surface through L_2. Gas masses with subcritical velocities are therefore trapped at this surface, eventually forming an almost spherical envelope.

In the case of β Lyr the bulkier of the two components is expanding. In other systems, e.g., U Cep and U Sge, the situation is reversed: the larger star is the one with the smaller mass. This is quite natural, since expansion increases the star's radius up to the Roche limit.

In the β Lyr system the large component is of spectral type cB9; this is a white supergiant of mass $m_1 = 63.2\odot$. The second component, according to Dadaev /17/, is an O star with weak lines (the lines may be diffused by the scattering of light off the electrons in the rarefied envelope of the star) and its mass is $m_2 = 42.1\odot$. Besides the cB9 lines, the spectrum of β Lyr shows lines previously assigned to a B5 star; according to O.Struve's interpretation, however, these are the lines of the expanding gaseous nebula which surrounds the system. At $\theta = 2\pi \cdot 0^p.08$, this nebula has a condensation which is steadily replenished by a stream of gas from the cB9 star. The lines of this stream in projection on the cB9 star (the satellite lines) appear at $\theta = 2\pi 0^p.008$. They give an average streaming velocity of -300 km/sec. This is slightly higher than the parabolic velocity. Therefore a large part of the gas is dissipated and only the fastest jets and condensations manage to circumnavigate the companion and return to the cB9 star with a velocity of +200 km/sec. Right before the central eclipse, for θ between $2\pi\ 0^p.925$ and $2\pi\ 0^p.980$, the satellite lines appear again, though they are weaker than at $2\pi \cdot 0^p.008$ (violet satellites). A schematic diagram of gas ejection is shown in Figure 38.

In the U Sge system gas is also ejected with high velocities: they reach 200 km/sec, which is clearly a supercritical value (greater than the parabolic velocity). In the U Cep system the exact streaming velocity is not known, but it is apparently subcritical. In this case, we are apparently dealing with continued flow of gas from the atmosphere of a gG2 subgiant to the atmosphere of the main B8 star.

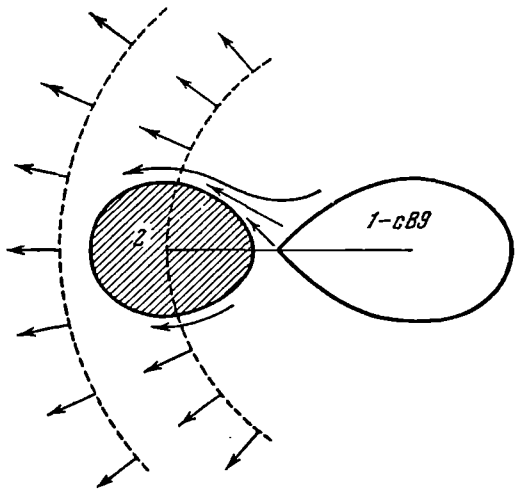


FIGURE 38

§37. Classification of close pairs. Eclipsing variables were traditionally divided into three classes: 1) Algol-type stars, 2) β Lyr stars, and 3) W UMa stars. This classification was based on a single phenomenological feature — the shape of the light curve. β Lyr stars and W UMa stars have a variable maximum phase (in distinction from Algol-type stars), this being a result of a highly pronounced ellipsoidal effect. The minima in β Lyr stars are of different depths, whereas the W UMa stars have almost equal minima. This approach clearly combines stars with widely differing characteristics under one category. Moreover, spectroscopic binaries, which are in effect close binaries not unlike the eclipsing binaries, do not fit into this classification.

The first comprehensive classification of binaries (including the close binaries) was proposed by Pike in 1931. Later, in 1944, Krat advanced a detailed classification of close binaries. This classification is based on physical characteristics. The stars are divided into seven classes. The leading classification parameter is the mass of the component stars. In the following, we list the seven classes in the order of decreasing (average) masses.

a) The two components are massive stars of spectral types O — B9. This class also includes binaries where both components are A-type supergiants (V 367 Cyg, and others). The average radius of the components is generally one order of magnitude greater than the radius of the Sun (⊙). Their masses are at least 4 — 6 solar masses. In extreme cases the masses exceed $m_\odot$ by one order of magnitude. The mean densities of the components are $10^{-2} - 10^{-3}\ \varrho_\odot$ ($\varrho_\odot$ being the mean density of the Sun) and less.

Since one of the component stars is often non-stationary, class a should be divided into two subclasses, aI and aII. Subclass aII comprises systems with non-stationary stars.

The light curves of subclass aI systems are generally highly regular. These are Algol-type systems. A typical representative of this subclass

is Y Cyg. GL Car is another typical system of this class, its components being stars of spectral types B3 and B4.

Light curves of subclass aII always show various irregularities and are highly skewed relative to the center line of the primary minimum. The figure of at least one of the stars is close to the Roche limit. Since the tidal deformations in this case are quite considerable, the light curve is generally of β Lyr type. A typical system of this subclass is β Lyr itself. A similar system is S Dor, where both components are non-stationary.

g) Class g systems are related to class a systems. In g systems one of the stars is a supergiant. In typical systems one of the components is a B-type star and the other is a red or yellow supergiant (ζ Aur), although often both stars are supergiants (ε Aur, VV Cep, etc.). Some of the g systems contain ordinary giants. The outermost envelopes of the supergiants and giants in these systems are very close to the Roche limit, so that these stars are non-stationary, although no clear indications of gas ejections have been observed. The g systems are characterized by long periods, from a few tens of days to tens of years. The bulkier star is generally the larger of the two (supergiant, giant).

b) The spectrum of the brighter star is between B0 and A9, whereas the fainter star (the satellite) is of types A0—F5. The periods of these systems range from a fraction of a day to a few days. The average masses and radii of the components are several times the radius and the mass of the Sun. The mean densities are a few tenths of the Sun's density. These systems are also divided into two subclasses.

Systems of subclass bI are stationary. The light curve is of Algol type. The ellipsoidal effect and the phase effect are inconspicuous. A typical representative of this subclass is AR Aur. Related systems are WW Aur and β Aur.

Subclass bII comprises the few non-stationary systems of class b. As in aII systems, the relative proximity of the components leads to a highly prominent ellipsoidal effect. The light curve is of β Lyr type. The gas ejection effect is inconspicuous and can be detected only by careful inspection, mainly from the inequality of the maxima (the periastron effect). Typical bII systems are V 505 Sgr and u Her.

c) One or sometimes both stars are subgiants. The period is generally more than 1 day. The component masses are close to the solar mass or about 2—3 solar masses. (This applies almost exclusively to the primary.) The radius of the primary is generally double the solar radius, whereas the radius of the satellite is several times the radius of the Sun. The mean densities of the subgiants are about $10^{-2}—10^{-3}\varrho_{\odot}$. The light curves are almost invariably of Algol type, since the primary is fairly small and the ellipsoidal effect is inconspicuous (remember that the shape of the light curve between eclipses is determined by the ellipsoidal effect).

Class c is also subdivided into stationary and non-stationary systems.

Subclass cI contains the few stationary systems whose components are fairly distant from the Roche limit. A typical system of this subclass is Z Her.

Subclass cII is characterized by gas ejection from the subgiant, whose surface in this case touches the Roche limit or lies sufficiently close to the Roche limit so that the star's chromosphere actually reaches the limit.

Although the existence of gas jets leaves no doubt, they are observed spectroscopically fairly seldom, mainly from various anomalies in the orbital elements (U Cep). The part of the light curve between the eclipses is distorted only in few systems. Skewness of the primary minimum is more often observed (U Cep, RS CVn, RW Tau, and others).

Typical cII systems are Algol, δ Lib, Z Dra, X Tri, RS Vul, TW Dra, RW Tau.

d) Both components are main-sequence stars. The satellite spectrum is F5, the primary has spectra from A0 to F0. The periods fall between wide limits. The satellite mass is close to that of the Sun. The mass of the primary star is generally at most double that of the Sun. The mean radius and density of the satellite, and sometimes of both stars, are almost those of the Sun. All the d systems are stationary. Since the surfaces of the components are far from the Roche limit (their relative dimensions are small), the phase and ellipsoidal effects are insignificant. Only few d systems are known; this is apparently due to their lower luminosity compared to systems of the previously described types. The mass of the satellite is not easily determined in this case, since the spectrum of the system shows only the lines of the primary. α CrB is a typical d system.

e) A small class with few systems. Both components are main-sequence stars of spectral types G—M. Period less than 1 day. The component masses are close to that of the Sun, their mean densities generally exceed the solar density. Systems of this class fall into two subgroups, stationary and non-stationary.

Subclass eI contains the few stationary systems, with YY Gem as a prototype. The light curves are of Algol type, since the relative dimensions of the stars are small.

Subclass eII comprises all those systems where at least one of the components touches the Roche limit. The light curves of these systems are of W UMa type. Gas ejection is inconspicuous, but it is nevertheless reflected in the light curve in the form of unequal maxima and a certain skewness of the minima. W UMa itself is a typical representative of this subclass. The component stars touch the Roche limit.

f) Both components in f systems are subdwarfs. This class apparently contains a great number of systems with periods invariably less than 1 day. The masses do not exceed the Sun's mass and are generally much smaller. The mean density of these stars is several times (and sometimes more) the mean density of the Sun. This class is also divided into two subclasses.

Subclass fI only comprises two systems, UX UMa and DQ Her = Nova 1934 Her. The components are physical variables of small amplitude and small light-variation period. In the DQ Her system and possibly in UX UMa, one or both components flare up as nova-like stars.

Subclass fII apparently comprises most W UMa eclipsing variables. Typical representatives of this subclass are AK Her, AG Vir, and RW CrB. The components are relatively early-type stars (A, F) of low mass (a few tenths of the Sun's mass). In the absence of spectroscopic observations, relatively high mean densities of the components are a sign of an fII system.

Table 1 lists the elements of systems representative of the various classes. The period is given to third significant digit. M_1 and M_2 are the bolometric stellar magnitudes, ϱ_1 and ϱ_2 are the mean densities of the components. A significant feature of this classification is that it covers all

TABLE 1

Star	Class	P, days	Spectral type		a_1		a_2		i, deg.	L_1	$\odot=1$		ϱ_1	ϱ_2	M_1	M_2	Source
			1	2		$\odot=1$		$\odot=1$			m_1	m_2					
YY Cyg	aI	2.996	O9.5	O9.5	0.208	5.9	0.202	5.9	88	0.59	17.4	17.2	0.085	0.084	−6.0	−6.0	Dugan. Princ. Contr. 12, 1931
β Lyr	aII	12.908	cB9	[O]	0.516	60.5	0.386	45.3	90.0	84	63.2	42.1	$4\cdot10^{-4}$	$5\cdot10^{-4}$	—	—	Dadaev. Bull. Pulkovo Obs., No. 152. 1954.
ζ Aur	g	972.15	K4	B8	0.169	201	0.0057	6.8	83.3	0.535	15.6	18.5	$1.93\cdot10^{-6}$	0.027	−4.6	−3.2	Krat. Bull. Eng. Obs., No. 19. 1937.
AR Aur	bI	4.135	B9	[A0]	0.0983	1.82	0.0983	1.82	88.5	0.537	2.55	2.30	0.43	0.38	0.3	0,6	Huffer, Eggen. ApJ, 106, 106, 1947
u Her	bII	2.051	B3	[B9]	0.287	4.28	0.276	4.11	77.3	0.782	7.79	2.76	0.114	0.029	−4.0	−2.6	Goldberg-Rogozinskaya. Bull. Pulkovo Obs., No. 147. 1951
Z Her	cI	3.993	F2	[gG1]	0.12	1.8	0.185	2.8	83.5	0.41	1.5	1.3	0.26	0.059	2.4	2.7	Baglow, Popper. ApJ, 106, 431. 1947
TW Dra	cII	2.807	A6	[gG5]	0.21	2.5	0.34	4.1	83.8	0.22	2.1	0.84	0 13	0.013	1.0	2.5	Bolokadze. Peremennye Zvezdy, Vol. 11, No. 5:375. 1958.
α CrB	d	17.360	A0	[G6]	0.069	2.9	0.021	0.87	88.3	0.989	2.5	0.89	0.10	1.31	−0.1	5.4	Kron. ApJ, 118:55. 1953.
YY Gem	eI	0.814	M1	M1	0.156	0.62	0.156	0.62	86.3	0.50	0.64	0.64	2.7	2.7	7.6	7.6	Kron. ApJ, 115:301. 1952.
W UMa	eII	0.334	F8	F7	0.441	1.11	0.312	0.79	78	0.63	1.30	0.65	0.94	1.24	4.1	4.7	Z. Kopal, M. Shapley. Jodrell Bank Ann., Vol. 1, f. 4, 1956.
UX UMa	fI	0.197	B3	A	0.23	0.50	0.17	0.37	85.2	0.96	1.20	1.20	9.0	24.0	—	—	Kalchaev. Bull. Pulkovo Obs., No. 155. 1956.
AK Her	fII	0.422	F9	G2	0.53	0.63	0.48	0.53	61.9	0.70	0.46	0.23	20	1.5	—	—	Ditto

the groups of the spectrum—luminosity diagram (except the white dwarfs). There is not a single known system which cannot be assigned to one of these classes. Moreover, in view of the great similarity in the properties of systems from the same class, the individual characteristics of systems in theoretical and statistical studies often can be replaced by the mean characteristics of the class. Finally, this classification presents in a condensed form the overall frequency of spectra, periods, and masses for the close binaries.

A significant shortcoming (unavoidable for any classification based on physical characteristics) is that poorly studied binaries sometimes cannot be classified.

In 1955 Kopal proposed a new classification, in which the close binaries are divided into three groups:

1) systems with detached components,

2) systems with semidetached components,

3) contact systems.

The first group includes all the binaries whose components lie virtually on the main sequence. The first (heavier) star is always the larger and has an earlier spectral type. Both components lie well inside the Roche limit (as defined by the point L_1).

The second group comprises binaries with only one main-sequence star. The primaries obey all the statistical laws characteristic of main-sequence stars. The satellites, in Kopal's opinion, have masses which are too small for their luminosities. The primary (heavier) component is generally the smaller of the two and has an earlier spectral type. The satellite's atmosphere generally touches the Roche limit.

The third group comprises W UMa and β Lyr stars (although β Lyr itself is not included). The two components lie not far from the main sequence. Their spectra fall between late A and early K. They do not obey the mass—luminosity relation. The heavier component is the larger of the two; it is generally of a later spectral type than its satellite and often of lower luminosity. The components almost fill the Roche limit. In Kopal's opinion, both components are probably in contact at point L_1, exchanging gaseous matter.

Kopal's classification is such that many of the known binaries simply do not fit in. These exceptions definitely include all the aII, g, eII, and fI stars (in Krat's classification). Moreover, generally it is impossible to determine whether or not the star's atmosphere touches the Roche limit, since gas ejection is observed only for very few binaries (if this conclusion is to be drawn from geometrical considerations, the relative elements are required with exceedingly high accuracy, again an improbable situation). Also note that numerous stars can be assigned simultaneously to two of Kopal's groups. Outstanding examples are V Pup (groups 1 and 3), AO Cas (groups 1 and 3), RT Lac (groups 2 and 3), and others.

§38. **Some remarks on stellar evolution in close binaries.** The considerable achievements of empirico-statistical work in stellar astronomy, which culminated in the construction of the spectrum—luminosity diagram and the discovery of the mass—luminosity relation and the period—luminosity relation for cepheids, led many observers to search for correlations between various elements of close binaries. These attempts, however,

failed entirely. The observational material obtained for close binaries only provided more detailed information on the spectrum—luminosity diagram and helped to improve the parameters of the mass—luminosity relation. The statistical studies of close binaries revealed absolutely no new regularities. It was established, however, that the existing material on close binaries is greatly influenced by numerous selection factors.

The detection probability of a spectroscopic binary depends primarily on the angle i (the K are proportional to sin i), on the mass ratio of the components, and on the period. If the mass of the primary is several times the mass of the companion, whose lines are not observed, K will be several times less than for a system with components of equal mass. For large P the distance between the stars is found to be so large that K often drops down to the threshold of accessibility of radial velocity measurements. Furthermore, the variation of K with time is inconspicuous in this case. Therefore for large P (if the components are not separated visually), only systems of classes a and g can be observed as spectroscopic binaries.

The situation is more complicated with regard to eclipsing variables. Here the significance of the angle i is much more prominent, since for i far from 90° no eclipse is observed altogether or else the light variation amplitude remains within the margin of error of observations. The relative sizes of the components (r_1, r_2 or a_1, a_2) also have a considerable influence on the probability of detection. The smaller these parameters, the lower is the eclipse probability. Large P also greatly reduce the detection probability, since the minima occur seldom and may easily escape the observer's notice.

Finally, the observed frequency of closed binaries of various classes is affected by another intractable factor, namely the nonuniform scanning of the sky areas by precision photometric techniques. Some sky areas (especially in the Southern Hemisphere) have been poorly studied.

Some statistical research was done in order to establish a possible relation between the period and the spectrum of close binaries, by analogy with the corresponding relation for the physical variables. One of the most comprehensive studies in this direction was carried out by Martynov /21/. His period—spectrum diagram, however, failed to reveal any correlation, but stars of different spectral types nevertheless were found to possess a certain lower-bound value of the period. This is a reflection of the fact that the distance between the components in a close binary cannot be arbitrarily small (definitely not less than the sum of the radii of the two stars).

Thus the available material on close binaries does not lead to any general (not even qualitative) conclusions concerning stellar evolution. We are therefore forced to abandon the large-scale statistical approach and to focus our attention on individual systems, especially the non-stationary specimens such as β Lyr, U Cep, and others.

First let us consider the evolution of orbits of close binaries. The variation of the period of a close binary is particularly significant for evolutionary considerations. Apart from the simple fact that the variation of the period can be determined with high precision from the observed epochs of minima, we are unable to establish any genetic relationship between close binaries with components from different classes of the spectrum—luminosity diagram without going into the question of the possible

evolution of a binary with a large P into a binary with a small P and vice versa.

The period of a close binary may change for the following reasons:

1) tidal and radial friction, and

2) loss of mass by the component stars.

Jeans has originally shown that the tidal friction, arising when the axial rotation period of a star is not equal to its orbital period, cannot substantially alter the period of a close binary, since the spin angular momentum of the components always constitutes but a minor fraction (not more than a few percent) of the total angular momentum of the system. The redistribution of angular momenta in the system in the result of tidal friction therefore cannot substantially increase the orbital momentum (which determines the period P). The radial friction — an effect theoretically predicted by Krat in 1937 — is simply incapable of increasing the period. Radial friction is an outcome of the variable deformation of stars moving in elliptical orbits /12/. It gradually increases the semiminor axis of the orbit and reduces the eccentricity. Any considerations involving a variation in the eccentricity of the orbits of close binaries should therefore be used with great caution. Tidal friction increases the eccentricity, while the radial friction tends to reduce it to zero.

The loss of mass by the stars is apparently the only cosmogonically significant factor capable of changing the period P.

Let us consider two different cases in which the variation of mass of one of the components leads to a change in P: 1) gas flows from the atmosphere of one component to the atmosphere of the other component, and the total angular momentum μ is conserved, 2) a gas stream is ejected with a supercritical velocity from the atmosphere of one of the components.

If the spin angular momenta associated with the relative rotation of the stars (relative to the line of centers) are ignored, μ can be written in the form

$$\mu = \frac{m_1 m_2}{m_1 + m_2} A v. \tag{141}$$

where A is the mean radius of the relative orbit and v is the mean velocity (a nearly circular orbit is assumed). Since

$$v = \frac{2\pi A}{P}$$

and

$$\frac{A^3}{P^2} \sim m_1 + m_2, \tag{142}$$

we have

$$\mu \sim 2\pi\alpha (1-\alpha) (m_1 + m_2)^{\frac{5}{3}} P^{\frac{1}{3}}, \tag{143}$$

where

$$\alpha = \frac{m_2}{m_1 + m_2}.$$

Taking the logarithmic derivative of (143), we get

$$\frac{d\mu}{\mu} = \frac{5}{3}\frac{d(m_1+m_2)}{m_1+m_2} + \frac{1}{3}\frac{dP}{P} + \frac{1-2\alpha}{\alpha(1-\alpha)}d\alpha. \tag{144}$$

In case 1, $\frac{d\mu}{\mu} = 0$ and $\frac{d(m_1+m_2)}{m_1+m_2} = 0$. Then

$$\frac{dP}{P} = -3\frac{1-2\alpha}{\alpha(1-\alpha)}d\alpha. \tag{145}$$

For $\alpha > 0.5$, dP has the same sign as $d\alpha$, and for $\alpha < 0.5$ it has the opposite sign. Therefore, when mass flows from m_1 to m_2, P increases if $\alpha > 0.5$ and decreases if $\alpha < 0.5$. In β Lyr, $\alpha = 0.4$ and P is therefore expected to decrease. The observations, however, give a different result. The P of β Lyr increases by $0^d.3838 \cdot 10^{-5}$ per period. The annual increment $\frac{\Delta P}{P}$ is $0.853 \cdot 10^{-5}$. This effect can be correctly understood if we remember that gas is ejected with parabolic velocity. In the limit, the star of mass m_1 can be treated as a point mass. Then for the change in the dimensions of the orbit we get

$$\frac{d \ln A}{dt} = -\frac{d \ln (m_1+m_2)}{dt}, \tag{146}$$

and for the change in the period

$$\frac{d \ln P}{dt} = -2\frac{d \ln (m_1+m_2)}{dt}. \tag{147}$$

Using (147) we obtain for the annual loss of mass

$$\frac{\Delta m}{m_1+m_2} = -0.427 \cdot 10^{-5}.$$

Hence it follows that the ejection of gas has been going on for no more than 10^5 years, since the present-day mass of the cB9 star is actually one of the largest among stars. If before expansion the cB9 star went through an equilibrium stage, this stage was very brief, since we know of no equilibrium stars of such large mass. It therefore seems that 10^5 years is the upper-bound value for the age of β Lyr. A similar conclusion is obtained for S Dor. In other systems, however, gas ejection provides no indication of age. We can only estimate the duration of the expansion stage responsible for the gas streams.

Since gas flow from one star to its companion invariably leads to the expansion of the latter star /15/, stellar masses in fact change only by the mechanism described under case 2 above (Eq. (147)). In close binaries, the period may only increase on the cosmic time scale. Therefore, a system with a large period will never evolve to a system with a small period.

If no gas ejection is observed, the star retains a constant radius. Moreover, we can maintain that these stars have not experienced any previous substantial compression, since generally the figures of the stars in close binaries are slightly less than the Roche limit, whereas the formation of a binary from any diffuse medium is bounded by the Roche limit. This is one of the main arguments in favor of the simultaneous formation of the two components from the diffuse medium.

If the mass of the components decreases in the course of evolution, whereas the stars remain on the main sequence (moving along the main sequence in the spectrum—luminosity diagram toward higher stellar magnitudes), low-mass stars can hardly be expected to occur in close binaries. We do not know but that eventually the components of β Lyr will become Sun-type stars with a much longer orbital period (such systems would be difficult to observe owing to the marked contraction in the relative dimensions of the stars), but it seems obvious that stars in binaries of subclasses bII, cII, eII, and fII have never had masses greater than their present-day mass, since their current figures as such are close to the Roche limit. In the past, these stars could only have been smaller. Moreover, if the loss of mass by evolving stars were monotonic, proceeding at a higher rate for high-mass stars, a definite levelling of the component masses would be observed in close binaries. This mass levelling would be more pronounced in c, d, e, and f systems. In fact, however, the mass ratio in these systems is on the whole the same as in a, g, and b systems.

These data are incompatible with the hypothesis according to which the stars evolve along the main sequence. Despite the overall paucity of the observational material, it seems evident that the main-sequence stars may form from the start in different parts of the main sequence. The main sequence is thus not an evolutionary sequence, and different main-sequence stars may be of one age.

It is apparently not by chance that the non-stationary star or the supergiant (expanded star) in a and g systems almost always has the larger mass. Indeed, if we remember that the low-mass component evolves at a slower rate than the high-mass star (it burns energy more slowly and the chemical composition changes slowly), a stationary star of low mass will take more time before it starts expanding (following the gradual increase in temperature with conversion of hydrogen into helium and increase of helium content). The mass of the star as if determines the rate of its evolution. The existence of perfectly stationary a and b systems suggests that even high-mass stars (Y Cyg, and others) initially pass through a stationary state. The non-stationary stars — Be stars, Wolf—Rayet stars, supergiants, and apparently some of the giants — probably constitute a second stage in the evolution of massive O and B stars, since the stationary and the non-stationary star (or the supergiant) in the system are of the same age and the observed difference between them is traceable to the difference in their masses. In binaries with equal component masses, both stars are either stationary (Y Cyg) or non-stationary (S Dor).

The longer duration of the stationary stage for low-mass stars (b and c systems) is apparently responsible for the great number of stars in spectral types A and late B in the spectrum—luminosity diagram. Here the transition to the giant phase is not as apparent as for the massive stars, although giants as a rule combine in pairs with B and A stars (TV Cas, SX Cas, U CrB, RV Tel, and others). Even more often, the second and generally non-stationary star is a subgiant (U Cep, U Sge, and others). Since the masses of subgiants are not greater than the solar mass, we have here a case of non-stationary low-mass stars. Non-stability may thus apparently set in during the early evolutionary stages too, since the primary star in a cII system is evidently a young star (from stellar kinematics and dynamics,

B and A stars are on the average younger than the main-sequence stars of later spectral types).

The existence of close binaries with two main-sequence stars (d system) whose spectra are of a later type than the spectra of the primaries in cI and cII systems can be understood if we assume that the two stars in these systems remained stationary for a long time, retaining their original mass and luminosity.

Stars of class f deserve particular attention. There is not a single f binary where a subdwarf combines with a main-sequence star. Subdwarfs combine only with subdwarfs. Either this shows that there is no genetic relationship between subdwarfs and other stars, or else it is evidence of the great age of subdwarfs compared to other stars. In this case, f systems could be regarded as "old" systems. Note that gas ejection is observed almost in all the f systems, which points to a common tendency of the atmospheres of subdwarfs to expand. Theoretically this is quite possible if we assume that the reserve of nuclear "fuel" in subdwarfs has been mostly depleted. When the last of hydrogen is converted into helium, the opacity of the stellar gas increases, the temperature and pressure at the center increase, and the star as a whole expands.

These are the only conclusions which can be drawn at this stage from direct observational evidence.

REFERENCES TO CHAPTER V

1. ZVEREV, M.S., B.V. KUKARKIN, D.Ya. MARTYNOV, P.P. PARENAGO, N.F. FLORYA, and V.P. TSESEVICH.— In: "Peremennye zvezdy," Vol.III, Gostekhizdat. 1947.
2. RUSSELL, H.N.— ApJ, 35:340. 1912; 36:54. 1912; 36:133. 1912.
3. KRAT, V.A.— Astron. Zh., 11:407. 1934; 12:21. 1935; 13:521. 1936.
4. SEMENOVA, N.N.— Bull. Pulkovo Obs., No.159. 1958.
5. MERRILL, J.E.— Contr. Princ. Univ. Obs., No.23. 1950.
6. SITTERLY, B.W.— Contr. Princ. Univ. Obs., 11, 1930.
7. MERRILL, J.E.— Contr. Princ. Univ. Obs., No.24, 1953.
8. KRAT, V.— Circular Pulkovo Obs., No.32. 1941.
9. KRAT, V. and S. PETROV.— Bull. Pulkovo Obs., No.140. 1947.
10. PETLAAR, J.— Rech. Astr. Obs. Utrecht IX, p.1. 1923.
11. PIOTROVSKI, S.L.— ApJ 108:36. 1948.
12. KOPAL, Z.— ApJ, 108:46. 1948.
13. WYSE, A.B.— Lick. Obs. Bull., No.496. 1939.
14. KRAT, V.A.— Astron. Zh., 21:154. 1944; Peremennye Zvezdy, 6:160. 1947.
15. KRAT, V.A. Figury ravnovesiya nebesnykh tel (Equilibrium Figures of Celestial Bodies).— Gostekhizdat. 1950.
16. KRAT, V.A.— Circular Pulkovo Obs., No.29. 1940.
17. KADAEV, A.N.— Bull. Pulkovo Obs., No.147. 1951.
18. STRUVE, O. and SU-SHU HUANG.— Occ. Notes of the R.A.S. 3, No.19. 1957.
19. KRAT, V.A.— Bull. Pulkovo Obs., No.163. 1959.
20. DADAEV, A.N.— Bull. Pulkovo Obs., No.152. 1954.
21. MARTYNOV, D.Ya.— In: "Peremennye zvezdy," Vol.II. Gostekhizdat. 1939.

Part Three

VARIABLES AND NOVAE

V.G. GORBATSKII

Chapter VI

INTRODUCTION

Physical variables are stars whose light variation is caused by various interior processes, whereas optical variables are stars whose brightness varies due to some extrinsic factor, such as periodic eclipsing by a companion star. Optical variables were treated in Chapter V. In the present chapter we are specifically concerned with physical variables so that the use of the unqualified noun "variable" need not cause any confusion.

The first star whose light was observed to vary (with the exception of a few novae known since antiquity) was o Ceti, which was aptly called Mira, the Miraculous. During the 200 years from the discovery of light variation of o Ceti (1956) to the publication of Pigott's first catalogue of variables, only 12 stars of variable brightness were discovered. The improvement of observation techniques in the middle of the 19th century (Argelander) and mainly application of photographic methods (E. Pickering) lent a more extensive and systematic character to the research of variables. About a thousand variables were known at the beginning of the 20th century. With further improvement of observational techniques variables were being discovered in ever increasing numbers, and by 1959 the list of variables included some 25,000 stars.

The end of the 19th century saw the beginning of a large-scale study of the spectra of variables. In 1894 Belopol'skii first observed variation of radial velocities in δ Cep and η Aql, and thus provided a new tool for probing the nature of variables. The application of photoelectric methods enabled astronomers to record small and fast light fluctuations, and this considerably enriched our knowledge of different types of variables. During the last decade considerable progress has been achieved in the study of motions of variables and their distribution in space.

Generalization of observational data led to the establishment of a number of regular features common to different types of variables. These regularities are of the greatest significance in studying the structure and evolution of star systems. The nature of individual types of variables, however, is insufficiently known, and a tremendous number of facts are still without theoretical interpretation. A pressing need in a comprehensive theory of variables was first felt when a simple comparison of observational data for some types of variables shed a new light on the entire question of stellar energy sources /1/.

§39. **Methods of study of variables.** Variable stars are generally found by surveying photographs of a certain sky area taken at different times. Several methods of photographic comparison are available, the most useful of which are the following:

1. Comparison on a blink microscope. The two plates are viewed alternately, and the image of the variable star appears pulsating because of the difference in the image diameter on the two plates. The pulsating star is thus readily detected against the background of fixed magnitude stars.

2. Superimposing a negative on a positive print of a photograph taken at some other time; the images of variable stars appear encircled by a dark or a light fringe.

3. Comparison on a stereocomparator.

In some cases the stellar spectrum suggests that the star is a variable. Spectroscopic observations led to the discovery of some novae and stars with long-period light variation, whose spectra are highly characteristic.

A detailed description of the methods of hunting for variables will be found in specialized literature /2/.

The light curve, plotting the stellar magnitude as a function of time, is one of the main characteristics of variables. The magnitude estimates needed for the plotting of the light curve are obtained by comparing the apparent brightness of the star in question with that of comparison stars.

The magnitude of a star can be determined by a variety of techniques:

a) from naked-eye brightness estimates,

b) from visual-photometric estimates,

c) from photographic photometry,

d) with an electrophotometer.

Each of these different methods is described in Vol. I of this Course, and their actual applications are discussed in /2/. Note that although the photographic method is still the commonest, photoelectric observations acquire an ever increasing importance. Visual-photographic observations are much too tedious to compete with any success with photographic observations (of the same accuracy). Naked-eye estimates, although relatively crude, enjoy considerable popularity because of their extreme simplicity: they are ideally suited for the large-scale work done by amateur astronomers.

The light variation of stars is not monotonic: their magnitudes fluctuate between certain limits. The difference between the maximum and the minimum brightness of a star is called the light variation amplitude, and the time between two successive maxima (or minima) is called the light-variation cycle. In cases when the light variation of a star is fairly regular, the star's cycle is called its light-variation period.

If T_0 is the epoch of some maximum and P is the period, the epoch T of any maximum is obtained from the equation

$$T = T_0 + PE,$$

where E is the number of periods during the time $T - T_0$. T_0 is called the initial epoch; T_0 and P taken jointly constitute the linear elements of the variable.

To draw a tentative light curve, all the available magnitude estimates are plotted on a graph. (The observation epochs, as well as T and T_0, are generally expressed in Julian Days, J.D.).

A smooth curve is then drawn through the various points (or through the normal points corresponding to whole groups of observations). If the light variation is fairly fast, the observation epochs are reduced to the Sun's center to avoid effects associated with the finite velocity of light. The corresponding epochs are called heliocentric, and the index $\odot$ is generally used to identify these quantities. Special tables /3/ and nomograms /4/ are available for purposes of speedy conversion.

If a variable star has a constant period, it is generally characterized by an a v e r a g e light curve. To construct this average curve, all the observations are reduced to a common cycle. This reduction, however, requires exact knowledge of the phase of each observation. The phase of an observation is defined as the difference between the observation epoch and the nearest preceding epoch when the star had a certain known magnitude (it is generally chosen as the epoch of maximum or minimum brightness). The phase is normally expressed in fractions of the period. Observations of close phase are combined into normal points through which the average light curve is drawn.

The approximate values of the period and the light variation amplitude are determined from the light curve. Further observations are used to improve these results by a variety of graphic and analytical methods (see /2/).

The periods of variables are often subjected to fluctuations. These fluctuations emerge from the so-called $O - C$ diagrams, which plot the deviations of the observed epoch of maximum from the epoch calculated using the linear equation. If the $O - C$ differences as a function of time are represented by a curve, the star's period is apparently variable. If the results of numerous observations give a strictly periodic $O - C$ curve and this periodicity is also apparent in other characteristics of the star, the natural conclusion is that the light variation displays a dual periodicity. By now dual periodicity has been firmly established for few stars only, almost all of which are RR Lyrae stars.

Mostly the trend of the $O - C$ curve is ascribed to the so-called cumulative errors, which build up because of natural imperfections in the light variation mechanism of the star. Certain stars, however, occupy an intermediate position between those with strictly periodic light variation and those without any constant period. In the latter category, the variation of cycle length definitely cannot be associated with the effect of cumulative errors. It thus seems that the period fluctuations of p e r i o d i c variables are not random either: they are probably inherent in the mechanism of stellar variation. The actual reasons for period variation apparently will remain unclear until a comprehensive theory of variables has been developed.

Brightness measurements of variables at different effective wavelengths (multicolor photometry) constitute a fairly recent development. The results of these measurements are used in plotting monochromatic light curves, from which the time variation of the color index is inferred. The three-color U, B, V system (see Chapter II) is the most popular. The color—magnitude diagrams are widely used in the analysis of variables.

The spectra of variables are studied by essentially the same methods as the spectra of constant-magnitude star; the exposure times, however, vary with the phase of observation.

§40. **Designation and catalogues.** Every newly discovered variable is assigned a temporary symbol from the list in use in the particular country or observatory where the discovery was made. For example, the variables discovered in the USSR are designated by the three letters SVS (Soviet Variable Star) and a serial number; variables discovered at the Harvard College Observatory, USA, are designated H.V. followed by the serial number, etc.

When the star has been firmly confirmed as a variable, its temporary designation is replaced by a definitive one.

Historical reasons led to the development of the following system of variable designation: the first 9 variables discovered in a certain constellation are designated by the letters of the Latin alphabet from R to Z followed by the name of the constellation; later stars are designated RR, . . ., RZ, SS, . . ., SZ, and so on up to ZZ. When the number of variables started increasing at an alarming rate in 1907, an additional range of symbols was added, AA, . . ., AZ, . . ., QQ, . . ., QZ. This system is clearly sufficient for labeling 334 stars in each constellation. Further variables are simply designated V335, V336, and so on, followed by the name of the constellation. Novae are designated by the letter N followed by the year of eruption and the name of the constellation; later on, they are assigned a usual variable star symbol. When a star designated by a Greek letter reveals light variation (as is the case for δ Cep, β Lyr), its notation is not changed. The boundaries of the different constellation are fixed by the International Astronomical Union (IAU).

Although this notation system of variables is not ideal, it can hardly be replaced without running a great risk of confusion, as the corresponding symbols are widely used in the literature.

Currently the definitive symbols are assigned to new variables during the compilation of catalogues. Since 1946 general catalogues of variable stars are being compiled and published by the USSR Academy of Sciences and the Shternberg State Astronomical Institute. The first catalogue of variables was published in 1948. Nine supplements were published between 1949 and 1955, and the 2nd edition of the catalogue listing 14,708 variable stars (including eclipsing binaries) was published in 1958 /5/.

In addition to catalogues of confirmed variables, there are special catalogues of suspected variables. In 1950, the USSR Academy of Sciences and the Shternberg Astronomical Institute jointly published such a catalogue listing 8134 "possibles" /6/.

The general catalogue of variable stars gives the following information:

a) the coordinate of the star in the epoch 1900.0 and its precession;

b) stellar magnitudes at the maximum and the minimum;

c) type of light variation and spectral type;

d) epoch, period, and time of rise from minimum to maximum;

e) literature reference.

The catalogue naturally cannot give a complete bibliography on each variable star, and special bibliographical indices are therefore published. In 1934—1936 the Berlin Observatory published Prager's bibliography for the years 1915—1933 for all the variables in Andromeda—Ophiuchus /7/.

The great project was not completed at that time, and only in 1952 Schneller published an additional volume containing the bibliography for the years 1916—1950 on variable stars in Orion—Vulpecula. The bibliography for all the variables identified between 1930 and 1938 and previously omitted references were included in Vol. IV published in 1957. The bibliography on stars discovered prior to 1915 was published in 1918—1922.

The various results on variable stars are published in a range of stronomical publications. The Soviet journal Peremennye Zvezdy (Variable Stars) published since 1928 is exclusively devoted to the research of variables. It publishes theoretical papers concerned with stellar variation phenomena, reports of new variables, and results of observations of individual known variables.

Amateur astronomers make an invaluable contribution to the observation of variable stars. In some countries (England, France, New Zealand, Japan), the work of amateur astronomers is published in special bulletins. These publications mostly contain only the results of brightness observations and the ephemeris (the predicted epoch of maximum) of the variable.

§41. Classification of variables. So far no satisfactory and consistent theory of stellar variation has been developed for most types of variable stars. The existing classification of variables is therefore based on various external signs and as such is inevitably tentative. The character of light variation considered in conjunction with spectral features generally serves as a basis for classification. Kinematic characteristics and the spatial distribution of stars are sometimes also considered.

In terms of light variation, the variables are divided into regular, semiregular, and irregular. Regular variables can safely be classified proceeding from the shape of the light curve and the length of the period. The classification of semiregular stars is less certain, and the group of irregular variables contains stars of widely differing light curves. We thus clearly require a further characteristic for the classification of variables, and especially irregular variables. The star's luminosity provides the missing data. In terms of luminosity, all the variable stars fall into two groups.

I. Stars which retain a high luminosity during the entire light-variation cycle; these variables correspond to the giant branch on the Hertzsprung—Russell diagram.

II. Stars whose luminosity is medium or low during the entire cycle (or at least at the minima).

Each of these groups includes stars with similar light variation curves. Stars under II generally show abrupt, discontinuous changes of magnitude, whereas the light variation of most stars under I is smooth and gradual. The classification of variables according to their luminosity thus brings out groups of stars with common light curves and it is therefore physically meaningful. Schneller used this luminosity classification /8/ and it was also adopted in a number of reference book.

The classification that follows is based on light variation and spectral features. It largely coincides with the classification system adopted by most authors, although it lacks in detail and identifies only the firmly established types of variables. The list below enumerates the different subdivisions of Group I in the order of decreasing periodicity. A brief description of the relevant characteristics is given for each type.

The high-luminosity variables include the following types:

1. Cepheids. These variables are distinguished by their strictly regular light variation. δ Cep is regarded as the prototype of this class. In terms of the light-variation period the cepheids are divided into two subtypes:

a) Short-period cepheids or RR Lyrae stars, with light-variation periods of less than one day. These are stars of spectral types A—F which constitute the halo subsystem of the galaxy. A typical representative of this group is RR Lyr.

b) Long-period cepheids with periods ranging from 1 to 80 days and light-variation amplitudes from 1^m to 2^m. Their spectral type at the maximum is F2—F4 and at the minimum G or K. Stars with longer periods have a later spectral subtype at the minimum. The absorption lines in the cepheid spectra shift periodically in phase with the light variation. The long-period cepheids show a pronounced concentration toward the galactic plane. A typical representative of this group of η Aql.

2. β Canis Majoris stars constitute a relatively small group of variables. They are characterized by strict periodicity and small amplitude of light variation (less than $0^m.25$). The periods range from 3 to 6 hours. Radial velocities variable. Spectral type B (luminosity class III—IV).

3. Long-period variables. The light variation is periodic, but the periods fluctuate (by as much as 10% of the average period, and occasionally more). The periods of these variables range from 70 to 1300 days. The light variation amplitudes are generally very large, reaching 7^m; they vary from one cycle to the next, together with the shape of the light curve. Most stars are of spectral type Me, although some have Se and Ce spectra. Near the maximum the spectra of long-period variables show bright hydrogen emission lines, and near the minimum they display bright metal emission. Both the absorption and the emission lines show an insignificant shift from their normal position (10—15 km/sec); the shift is a function of phase. The long-period variables reveal but a slight concentration toward the galactic plane. o Cet is a typical representative of this group.

4. RV Tauri stars. Most of the time the light variation of RV Tauri stars is periodic, with periods ranging from 30 to 150 days. The minima are of varying depth, and the deeper (primary) minima generally alternate with shallower secondary dips. The amplitudes of these stars are between $0^m.8$ and $3^m.5$. Spectral types F, G, or K. Some stars acquire bright hydrogen lines near the maximum and TiO absorption bands near the minimum. The average brightness of some RV Tauri stars is also variable, with a period of a few years.

5. Semiregular and irregular late-type variables. These stars include:

a) "Yellow" semiregular stars (spectral types F, G, and K). Their light curves are on the whole similar to the RV Tauri curve, though somewhat less regular. S Vul is a typical representative.

b) "Red" semiregular stars (M type) with cyclic light variation. The length of each cycle is about 100 days and the amplitude is a few tenths of a stellar magnitude. Some of these stars have a constant average magnitude (e.g., AF Cyg, RS Cnc), whereas the average brightness of others is slowly variable (μ Cep).

c) Irregular late-type variables, without any periodicity in their light variation. The amplitudes are invariably small ($\leqslant 1^m$). Most of these stars have M-type spectra, although K, C, and S spectra are also observed. A typical representative is U Del.

6. Be stars. B-type stars with bright hydrogen emission lines in their spectra. Their magnitude is variable, but the light-variation amplitude is generally small ($0^m.1 - 0^m.2$); in some cases, however, it may reach $1^m.5$ (γ Cas).

Be stars have variable spectra, and are therefore classified as spectroscopic variables.

Spectroscopic variables are generally characterized by small light fluctuations, and are mainly detected from the pronounced variations in their spectra.

7. α^2 Canum Venaticorum stars. This fairly small group provides another example of spectroscopic variables. These are A stars with a strong magnetic fields. Their spectra show lines of rare-earth elements whose intensity varies with periods from 1 to 20 days. The light variation is most insignificant (with amplitudes $< 0^m.1$).

We will now consider the classification of medium- and low-luminosity variables. This group includes stars with tremendous light-variation amplitudes, which is indicatory of the catastrophic nature of processes responsible for light variation. For this reason these variables are sometimes described as "cataclysmic." They include the following types:

1. Novae. These are stars whose brightness increases rapidly (in a few days or weeks) by a factor of 10^4 (by $10^m - 11^m$). This brief "eruption" is followed by a gradual decay. The spectra of novae in the postmaximum stage show wide emission bands of hydrogen, helium, and other elements; the structure of these bands points to ejection of stellar matter with high velocities.

2. Supernovae. Similar to novae, but their brightness increases during the eruption by as much as 20^m.

3. Recurrent novae. These stars erupt like ordinary novae, though with a smaller amplitude ($7^m - 9^m$). The eruptions recur at an interval of a few decades or centuries. A typical representative of this group is T CrB.

Stars of the following group are closely related to the above types:

4. Z Andromedae stars or nova-like stars. These stars derive their name from the recurrent eruptions which are similar in many respects to the eruptions of novae. The brightness of the erupting star increases by $4^m - 5^m$, and the declining branch shows damped oscillations. Z And spectra present the combined features of a hot star surrounded by an envelope (hydrogen and helium emission lines) and of a cold star (TiO absorption bands).

The following stars are also often included among the nova-like variables.

5. U Geminorum stars, and

6. Z Camelopardalis stars.

A detailed study of the light curves and spectra of these stars shows, however, that they have very little in common with novae. U Gem stars show cyclic light variation. The height of the maximum varies from one

eruption to the next, and the brightness slightly fluctuates during the prolonged minimum. The light variation amplitude is a few stellar magnitudes. The spectrum between successive maxima shows strong hydrogen emission lines and continuum emission, whose profiles are definitely unlike those of the emission lines in the spectra of novae. At the maximum the spectrum shows very wide unshifted absorption lines. The light variation of Z Cam stars is not unlike that of U Gem stars, but occasionally the succession of maxima is interrupted for a fairly long period, during which the star retains a constant magnitude which lies between the maximum and the minimum.

Additional stars included in this group are the following:

7. R Coronae Borealis stars. Most of the time stars of this type retain a constant magnitude; occasionally the brightness falls abruptly a few stellar magnitudes, and subsequently the star waxes gradually to its usual brightness. The energy distribution in the spectrum invariably corresponds to G-type stars throughout the cycle. During the stage of constant magnitude, the star has an absorption spectrum but bright lines develop when the brightness falls. The luminosity of R CrB stars during the constant brightness stage is high, corresponding to the luminosity of giants; in minima these stars are typical dwarfs.

8. RW Aurigae stars. The spectral types of these stars range from B to M inclusive, but most stars are of spectral types F, G, and K. Their light variation is quite irregular; sometimes the brightness varies at a fast rate, changing by 3^m-4^m in less than one day, whereas in other cases the variation is much slower. No detailed classification has been developed so far for RW Aur stars; nevertheless T Tau stars are distinguished as a separate subtype. T Tau itself and the related stars have a distinctive yellow color (G—K spectra) with bright lines in their spectra. They show a peculiar distribution in space: all the known T Tau stars are concentrated in several dense groups, known as T associations. In terms of luminosity, RW Aur stars are main sequence stars and T Tau stars are dwarfs.

9. UV Ceti stars. Stars of very low luminosities and spectral type dMe. Their spectra show bright lines of hydrogen and ionized calcium. They are subject to sudden flares, the brightness of the star increasing in a few minutes (!) by several stellar magnitudes and then dropping back to its initial value in not more than one hour. The flare is accompanied by a continuous emission spectrum which veils the absorption spectrum.

REFERENCES TO CHAPTER VI

1. AMBARTSUMYAN, V.A. Continuous emission and sources of stellar energy.— Bull. Byurakan Obs., 13, Erevan. 1954.
2. ZVEREV, M.S., B.V. KUKARKIN, D.Ya. MARTYNOV, P.P. PARENAGO, N.F. FLORYA, and V.P. TSESEVICH. Metody izucheniya peremennykh zvezd (Methods of Study of Variable Stars).— Gostekhizdat. 1947.
3. PRAGER, R. Tabellen der Lichtgleichung. Kl. Veröffenlichungen. Berlin—Babelsberg, No.12. 1932.
4. ZVEREV, M.S.— Trudy, Shternberg Astron. Inst., 7, No.1. 1936.
5. KUKARKIN, B.V., P.P. PARENAGO, Yu.I. EFREMOV, and P.N. KHOLOPOV. Obshchii atlas peremennykh zvezd (A General Catalogue of Variable Stars), 2nd ed.— AN SSSR. 1958.

6. KUKARKIN, B. V., P. P. PARENAGO, Yu. I. EFREMOV, and P. N. KHOLOPOV. Katalog zvezd, zapodzorennykh v peremennosti (A Catalogue of Suspected Variables). — AN SSSR. 1951.
7. PRAGER, R. Geschichte und Literatur des Lichtwechsels der veränderlichen Sterne. 2. Ausgabe (I Band. Andromeda — Crux, Berlin, 1934; II Band. Cygnus — Ophiuchus, Berlin, 1936).
8. SCHNELLER, H. — Geschichte und Literatur des Lichtwechsels der veranderlichen Sterne. 2. Ausgabe (III Band. Orion — Vulpecula, Berlin, 1952; IV Band. Berlin, 1957).

GENERAL MONOGRAPHS

1. KUKARKIN, V. V. and P. P. PARENAGO. Fizicheskie peremennye zvezdy (Physical variables). — ONTI. 1937.
2. PAYNE GAPOSCHKIN, C. and S. GAPOSCHKIN. Variable Stars. — Harv. Obs. Monographs, No. 5. 1938.
3. CAMPBELL, L. and L. JACCHIA. The Story of Variable Stars. — Philadelphia. 1945.
4. VORONTSOV-VEL'YAMINOV, B. A. Gazovye tumannosti i novye zvezdy (Gaseous Nebulae and Novae). — AN SSSR. 1948.

O.A. MEL'NIKOV

Chapter VII

CEPHEIDS

§42. Long-period cepheids. These stars have been studied by numerous authors. In describing their characteristics, we will mention only few of them by name as typical cases.

1. Light curves of cepheids are characterized by the following "light elements": the time or the epoch of the maximum T_0 (initial time), the period P, and the mean $\overline{m}$ or median m_{med} magnitude. *

The variation of the cepheid magnitude over a short period of time can be expressed by a linear relation from which the ephemeris, i.e., the Julian Date of the forthcoming (or past) maxima, can be calculated:

$$T_{max} = T_0 + P \cdot E.$$

Here E is a whole number of periods (the number of epochs) elapsing since the epoch T_{max}. At any other time of observations, we have for T

$$T = T_0 + P \cdot E + \Phi^d,$$

* The median magnitude is $m_{med} = m = \frac{m_1 + m_2}{2}$, where m_1 and m_2 are the stellar magnitudes at the minimum and at the maximum, respectively. The mean magnitude $\overline{m}$, also used to express the parameters of period–luminosity and other curves, is a simple average of the available magnitude estimates. The available estimates are assumed to be uniformly distributed over the entire light curve. Alternatively, $\overline{m}$ can be defined as the magnitude dividing the area enclosed by the light curve into two equal parts, above and below the corresponding horizontal line. The median magnitude m_{med} is equal to the mean magnitude $\overline{m}$ only for stars with symmetric light curves, i.e., types c and C; it is greater than the mean magnitude (i.e., the median luminosity is lower) for stars with skewed light curves (types a, A and b, B, see below). For short-period cepheids the median magnitudes are virtually constant (with a small dispersion), a highly useful feature for the determination of distances to globular clusters (without previous knowledge of the light period). The mean magnitudes of the short-period stars, however, are not constant: they increase with increasing period (the stars grow weaker) because of the correlation between the shape of the light curve and the period (the Hertzsprung–Ludendorff relation). Currently the mean magnitudes of cepheids are calculated from the ratio l/l_0, and not from the equality $m = -2{,}5 \lg l/l_0 + m_0$, where l is the apparent (relative) luminosity, or intensity. Only this quantity characterizes the intrinsic radiation of the star. m_0 and l_0 are associated with the particular scales used. The magnitudes are used to plot the light curve in units of intensity $l/l_0 = 10^{-0.4(m-m_0)}$; it is a well-known fact that if $m_{min} - m_{max} > 1 - 2^m$, the light and intensity curves are entirely dissimilar, the best example being the long-period variable χ Cyg in Figure 39 (type D). After that, the area under the curve is measured planimetrically, the mean relative luminosity $\overline{l/l_0}$ bisecting this area is found, and it is used to determine the mean magnitude m. Clearly $m \neq \overline{m}$; in general m is also greater than $\overline{m}$, which corresponds to luminosity $l < \overline{l}$ (in this case we take $\overline{m}_0 = 0$, $l_0 = 1$).

where $(T_0+P\cdot E)$ is the Julian Date of the preceding maximum, and Φ^d is the phase in days. Alternatively,

$$\frac{T-T_0}{P}=E+\Phi^P.$$

Here Φ^P is the phase in fractions of the period.

The shape of the light curve of cepheids is characterized by its amplitude, skewness $\varepsilon=\frac{T_M-T_m}{P}$, where T_M is the time of the maximum, T_m the time of the minimum, and a range of other features, in particular the presence of a "hump," its phase and area (in fractions of the total area enclosed by the light curve), the Fourier components of the light curve, etc.

Figure 39 shows specimen photoelectric curves of the long-period cepheids δ Cep, η Aql, and ζ Gem (Guthnick, Weyl, and Eggen). δ Cep has a regular but skew light curve (the rise branch being steeper than the descending branch). ζ Gem, conversely, has an almost symmetric light curve. η Aql, on the other hand, has a "hump" on the descending branch of its skew light curve, near the phase 2^d.

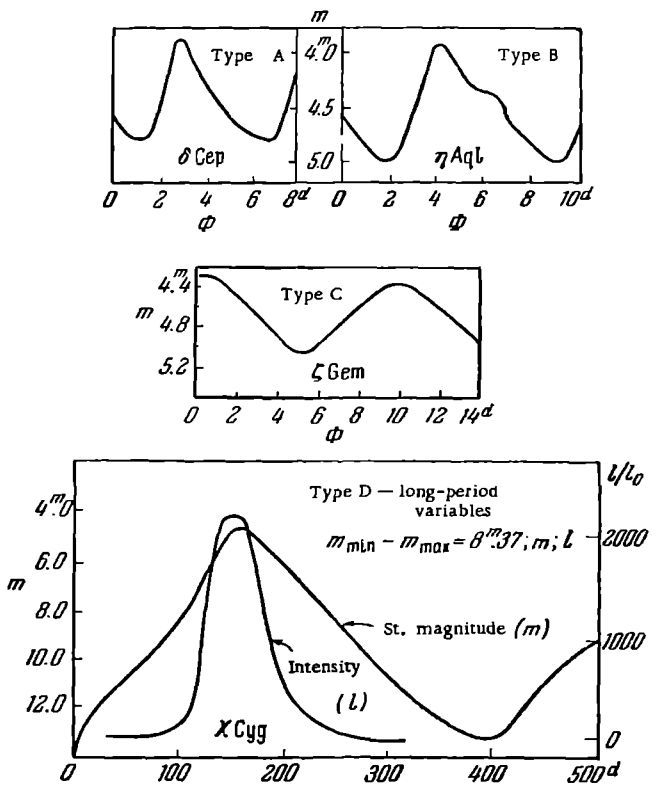


FIGURE 39

The peculiar features of the light curves of long-period cepheids necessitated the construction of "standard" light curves for classification

purposes; the standard light curves are based on the average results of individual observations using different instruments at different observatories and combine curves of similar periods, etc. These standard curves are furthermore free from certain errors and are thus more reliable than the individual curves. The most widely used are the standard curves of Kukarkin and Parenago (1936—1937). Some of these curves are shown in Figure 40.

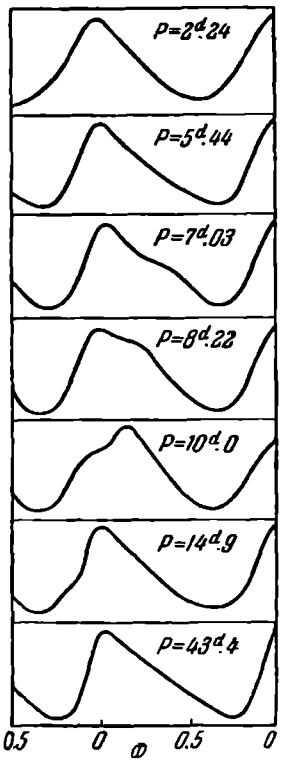


FIGURE 40

We see from the figure that cepheids with relatively short periods (2—7^d) have skew curves, similar to those of the cepheids with periods of 14^d and longer. Cepheids with periods of about 10 days have virtually symmetric curves ($\varepsilon \approx 0.5$). Stars with periods of 6 days have a secondary wave — a "hump" — on the descending branch. As the period increases, the hump climbs up the descending branch. At the same time its area, generally expressed in units of the area enclosed by the entire curve, increases. For cepheids with periods slightly less than 10 days the hump reaches the primary maximum and displaces it, so to say. The displaced maximum apparently migrates down the ascending branch in the form of a secondary, weaker wave. In cepheids of longer periods this secondary wave grows weaker and then disappears altogether. We are probably dealing with a superposition of two oscillations. Their relative contributions are apparently different for cepheids of different periods. The only exception are the 10-day cepheids, where both oscillations are roughly equivalent. This critical period is also manifested in other characteristics of the long-period cepheids; in particular, some of the critical-period cepheids, e.g., $\varkappa$ Pav ($P \approx 9^d$), show significant variation of the period (Kukarkin and Gits, 1937). For cepheids of longer and shorter periods, no significant variation of periods is observed. There are, however, some indications that the period of δ Cep slowly decreases (by about 0.1 sec annually), but the data available are insufficient for a conclusive proof of this variation (Sentsova, 1959).

The epochs of maximum of a number of stars, in particular of ζ Gem, cannot be adequately fitted with a linear relation. An additional quadratic term is therefore introduced

$$T_{\max} = T_0 + P \cdot E + qE^2.$$

Here is either positive or negative depending on the sign of period variation (whether incrementing or decrementing). Generally q is fairly small. Calculating $T_{\max}$ for the epochs E and $E + 1$ and subtracting one from the other, we obtain an expression for the period, i.e., for the time between two successive maxima:

$$T_{\max}^{E+1} - T_{\max}^{E} = P = P_0 + 2qE + q \approx P_0 + 2qE.$$

Hence it follows that the presence of a quadratic term makes the period roughly proportional to time, or E. The period may either decrease ($q < 0$) or increase ($q > 0$). Evidently,

$$\frac{dP}{dE} = 2q$$

is the time rate of change of the period. For ζ Gem, e.g., $2q = -1^d.15 \cdot 10^{-6}$; therefore $T_{max} = 2{,}410{,}639.801 + 10^d.153527\,E - 0^d.575 \cdot 10^{-6}E^2$. A number of astronomers, in particular Florya and Kukarkin (1932) and Kulikovskii (1935), have shown that the rate of change of the period $\left(2q = \frac{dP}{dE}\right)$ and the period (P) are related by the equality

$$\lg\left|\frac{1}{P}\,\frac{dP}{dE}\right| = -|a| + |b| \lg P.$$

This is the equation of secular variation of the cepheid periods.

A detailed study of this problem has shown, however, that the observed variation of the period may be due to evolutionary factors, random errors, light detector properties, insufficient "accuracy" of the variable (significant fluctuations in P), etc. The superposition of all these factors may produce an impression of a variable period (cumulative errors).

T. Sterne (1934) in Harvard developed a special method for isolating the contribution from the real, evolutionary variation of the period. Unfortunately the application of this method to a number of variables failed to provide a conclusive answer. Nevertheless, the role of cumulative errors must be established, as it is of bovious interest. Note that neither significant changes in the shape of the light curves nor secular variation of the mean magnitude were observed for the long-period cepheids. This is due, in particular, to the fact that high-precision photoelectric curves (accurate to within $\pm 0^m.01$) are a relatively recent achievement, whereas the search for secular variations requires long observation series. Besides the progressive variation, the periods may also experience irregular or even periodic fluctuations. In the latter case, the quadratic (parabolic) term in the equation of T_{max} should be replaced by a sine term of the form $A \sin(B \cdot E + C)$. Long-period cepheids, however, generally fail to reveal fluctuations of this type.

TABLE 1

λ (Å)	δ Cep	η Aql
3530	$1^m.48$	$1^m.45$
4220	1 .41	1 .32
4880	1 .08	1 .00
5700	0 .86	0 .76
7190	0 .61	0 .57
10300	0 .43	0 .37

Note that the light curves of long-period cepheids in extragalactic nebulae, e.g., M 31 (in Andromeda), M 33 (in Triangulum), and NGC 6822 (in Sagittarius), are similar to the light curves of the galactic cepheids.

The light variation amplitudes A_λ of the cepheids decrease with increasing wavelength. Monochromatic light amplitudes of δ Cep and η Aql obtained by photoelectric methods (Stebbins, Whitford, et al.) are listed in Table 1.

If the energy distribution in the continuous spectrum of cepheids can be represented by Planck's equation with a hypothetical, so-called spectrophotometric or color temperature T_s, the variation of amplitude with

wavelength is described by

$$A_\lambda = -2.5 \lg \frac{R_m^2 J(\lambda T_m)}{R_M^2 J(\lambda T_M)} = -5 \lg \frac{R_m}{R_M} - 2.5 \lg \frac{e^{\frac{c_2}{\lambda T_M}} - 1}{e^{\frac{c_2}{\lambda T_m}} - 1} =$$

$$= -5 \lg \frac{R_m}{R_M} + \frac{1.086}{\lambda} \left(\frac{c_2}{T_m} - \frac{c_2}{T_M} \right) + \Delta x(\lambda, T_m, T_M).$$

Here R is the radius of the star and

$$\Delta x(\lambda, T_m, T_M) = -2.5 \lg \frac{(1 - e^{-\frac{c_2}{\lambda T_M}})}{(1 - e^{-\frac{c_2}{\lambda T_m}})},$$

the subscripts M and m referring to the maximum and the minimum, respectively. For long-period cepheids of moderate temperature, Δx is generally small. In the infrared, i.e., for $\lambda \to \infty$,

$$A_\lambda \to -5 \lg \frac{R_m}{R_M} - 2.5 \frac{T_m}{T_M} = \text{const}.$$

In the ultraviolet, for $\lambda \to 0$, $A_\lambda \to \infty$. It is clear, however, that the limiting values are applicable only if the energy distribution in the entire spectrum is Planckian. Observations of the Sun, which is close in its spectral type to average-period cepheids, show that this is not quite so.

In 1936, A. Bleksley established for cepheids the relation $R \cdot T = \text{const}$ or approximately $R \cdot T_s = \text{const}$. If at long wavelengths we use the Rayleigh—Jeans formula $\left(J(\lambda, T) \sim \frac{T}{\lambda^4}\right)$, we have for the light amplitude

$$A_\lambda = -5 \lg \frac{R_m}{R_M} - 2.5 \lg \frac{T_m}{T_M} = -2.5 \lg \frac{T_M}{T_m}.$$

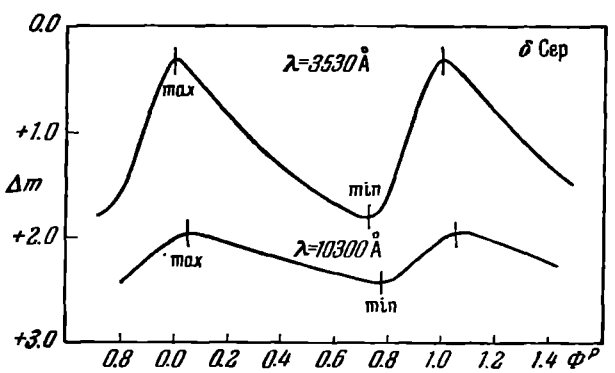


FIGURE 41

Thus, in the far infrared, the light variation amplitude reverses its sign and the curve is inverted. The radiation of a star decreases with increasing temperature, and vice versa. There is a certain tendency to a displacement of the phase of the maximum. This is evident from the light curves of δ Cep (Figure 41). On passing from the ultraviolet to the infrared region, a

certain lag in the phases of maxima and minima is observed. The corresponding epochs of the infrared curve are somewhat later (photoelectric observations of Stebbins and others). This problem was investigated by T. Dambara in 1951. Unfortunately, radiometric measurements of the light curves of δ Cep and η Aql at effective wavelengths of 20,000 Å = 2μ failed to reveal the expected lag of maximum and minimum phases. The accuracy of these observations was very low, however, much lower than the accuracy of the photoelectric curves. Kukarkin studied 23 long-period cepheids and derived an expression for the lag of the maximum phase of the visual curves relative to photographic curves, $\Delta\Phi = 0^P.079 - 0^P,1025 \lg P$. *

In the author's opinion, the apparent shift of maxima and minima between the short-wave and the long-wave curves is due to differences in the rate at which the photospheres of cepheids grow "transparent" as the star compresses and expands. Thus, if the photosphere expands near the minimum (and the density rapidly decreases), the observed minimum occurs earlier than the true minimum does. This is so because as the photosphere becomes transparent, the radiation reaches the observer from deeper lying strata even before the minimum has been reached and the star apparently contracts. A similar effect, though on a larger scale, is observed in the expanding photospheres of novae. The results are different at different wavelengths, since the opacity of the photosphere is different, which explains the phase lag. The expression for the light amplitude is briefly written in the form

$$A_\lambda = A_R + A_{T,\lambda}.$$

Here A_R and $A_{T,\lambda}$ are the changes in luminosity associated with changes in the radius and in the temperature of the star, respectively. Therefore, if we choose two points with equal color indices on the light curve, the difference in stellar magnitudes Δm_R will correspond to the contribution from the changes in the surface of the star between two given phases. Thus,

$$\Delta m_R = -2.5\Delta \lg (4\pi R^2).$$

Writing ΔR and ΔR_2 for the increment of the radius $R_i - \overline{R} = \Delta R_i$ ($\overline{R}$ being the average for the phases of equal color, when the temperature amplitude is zero), we have

$$2.5 \lg \frac{(\overline{R}+\Delta R_1)^2}{(\overline{R}+\Delta R_2)^2} = \Delta m_R,$$

$$\left(\frac{\overline{R}+\Delta R_1}{\overline{R}+\Delta R_2}\right)^2 = \lg^{-1} \Delta m_R \cdot 0.4,$$

$$\frac{\overline{R}+\Delta R_1}{\overline{R}+\Delta R_2} = \lg^{-1} \Delta m_R \cdot 0.2 = n.$$

Finally, for the average radius we get

$$\overline{R} = \frac{n\Delta R_2 - \Delta R_1}{1-n}.$$

Here n is read off the light curve, and ΔR_1 and ΔR_2 from the radial velocity curves. We can thus calculate $\overline{R}$ (provided the phase shift of m and v_R is disregarded).

* From observations at various wavelengths for δ Cep (Stebbins), say, $\Delta\Phi = 0^P.000$, $0^P.009$, $0^P.024$, $0^P.037$, $0^P.050$ for $\lambda = 353$, 422, 488, 570, 719 and 1030 mμ.

Thus, Stebbins (1953) obtained from his photometric measurements* $\bar{R} \cdot 10^{-7}$ km = 3.7 ± 0.1, 4.7 ± 0.2 and 0.50 ± 0.06 for δ Cep and η Aql, respectively.

The above expression for $A(\lambda, T)$ can be applied with the least squares method to calculate (if the amplitudes A are known for various wavelengths λ) by successive approximations the values of T_M and T_m without reference to low-temperature ($T \approx 2600°$) terrestrial sources, such as lamps, arcs, etc. These reference standards are highly inadequate, since the transition from low to high temperatures involves a tremendous loss in accuracy (since the formulas are exponential). The results obtained by this new method for δ Cep and η Aql are listed in Table 2. **

TABLE 2

Star \ Temperature	Degrees Kelvin				Spectrum
	Becker (λ=557 mμ)			Mel'nikov (absolute λ=570 mμ)	
	11000°	13500°	18000°		
δ Cep					
T_M	6200°	6800°	7600°	7080°±870°	F5
T_m	5000°	5400°	5800°	5040°±480°	G3
η Aql					
T_M	5600°	6000°	6600°	6280°±1540°	F6
T_m	4200°	4400°	4800°	4760°± 980°	G3

Table 2 also lists the spectrophotometric temperatures T_s obtained by W. Becker in 1937—1940 following an extensive study of a number of cepheids using reference stars with various temperature zero points. T_s was determined by the conventional method, from the relative gradients of the cepheids

$$\Delta\varphi = \varphi_{Cep} - \varphi_0.$$

Here φ_0 are the absolute gradients of the reference stars. The absolute gradients of the cepheids can be found from

$$\varphi_{Cep} = \frac{c_2}{T_{Cep}} \left(1 - e^{-\frac{c_2}{\lambda T_{Cep}}}\right)^{-1}.$$

The uncertainty is associated with the ambiguity in the definition of φ_0 in various temperature scales ranging from 10,000 (Potsdam) to 18,000 °K

* Note that the determination of ΔR_i by integration of the radial velocity curve between points of equal color index is complicated by the relative phase shift of the monochromatic light curves (see previous footnote). For example, Sycheva's data (1949) based on Stebbins's six-color photometry give the following expression for the phase shift of the light curve as a function of wavelength:

$$\Delta ph_\lambda = -0^P.0408 + 0^P.00000714\lambda\ (\text{Å}).$$

This phase shift and the phase difference between light and velocity curves gives different average radii if different combinations of light curves are used. For example, the $U+G$ and $V+R$ curves (3530, 5700 Å and 4220, 7190 Å) give average radii of 21.5 and 36.4 $R_\odot$, respectively. Other color combinations give intermediate values. The problem is further complicated by the phase relations between the light curve and the radial velocity curve, which remain even if we ignore the famous 1/4 period shift (the "mirror image" shape of the two curves).

** At the 1960 Conference on Variable Stars in Rostov, V. V. Prokof'eva reported the results of a reduction of the monochromatic amplitudes of a number of long-period cepheids by Mel'nikov's method. The results were fully consistent with Mel'nikov's original data.

(Old Greenwich) for A0 stars. More exact determinations of T_{Cep} by the new method, without any zero-point uncertainty, were carried out for δ Cep and η Aql by Mel'nikov in 1955 and are also listed in Table 2. The weak point in this case is the application of blackbody or at least "gray body" equations to real cepheids. This is, however, not a bad approximation, since the Sun, as we know (whose spectrum is close to that of δ Cep and η Aql), satisfies the gray body hypothesis to first approximation.

The relative gradients of cepheids in relation to A0 stars can be found with fair accuracy. The determination of their spectrophotometric temperatures therefore provides a method for independent determination of the temperature of A0 stars. The spectrophotometric temperature of these stars was found to be $T_{A0} = 15{,}400$ °K. Analysis of the A_λ vs. P curve gives T_c as a function of P for the entire sequence. This result was originally obtained by Mel'nikov in 1944. At that time, however, no reliable A_λ values were available for several spectral regions and for numerous cepheids. The study was therefor based on two A_λ values only — photographic and visual; the results are thus highly tentative.

Mel'nikov originally explained the observed spectral characteristics of long-period cepheids using Kozyrev's and Chandrasekhar's theory of extended photospheres (1934). The fit was on the whole satisfactory and even shed light on the anomalous excitation and ionization in cepheid atmospheres.

Most interesting theoretical calculations of light amplitudes of cepheids using models of F—G stars were carried out and published by Canavaggia and Pecker in 1952. The ratio of maximum to minimum radius was found to be 1.2, which is in good agreement with the results of observations (1.1 according to Stebbins's six-color photometry).

The most detailed and fairly accurate photographic observations of amplitudes and magnitudes of various cepheids at all the wavelengths from red to violet were carried out by Becker (1937—1941). Eggen studied photoelectrically the magnitudes of various long-period cepheids in two spectral regions (32 stars) using the 12-in. Lick Observatory refractor. He classified all the stars into three groups, designated A, B, C. Typical representatives of these groups were δ Cep, η Aql, and ζ Gem, respectively. The main feature used in this classification was the dependence of the light and color amplitude on the period. A similar division, although based on the shape of the light curve only, was proposed by S. Bailey back in 1902 for short-period cepheids (RR Lyrae stars). The corresponding groups were marked a, b, c. Parenago studied in more detail the classification of cepheids (1955) and showed that the period—photographic amplitude relation (even for amplitudes measured photoelectrically) for long-period cepheids is more of a correlation than a dependence. The above classification of the long-period cepheids into groups A, B, and C is thus highly tentative, though not entirely improbable.

2. Spectrum variation curves. The spectra of cepheids, mainly studied at the Harvard Observatory, show that on the average the amplitude of spectrum variation from maximum to minimum corresponds to one spectral type. At the maximum or near it, the star is of an earlier spectral type than at the minimum. The spectra of δ Cep in various phases of the light curve are shown in Figure 42.

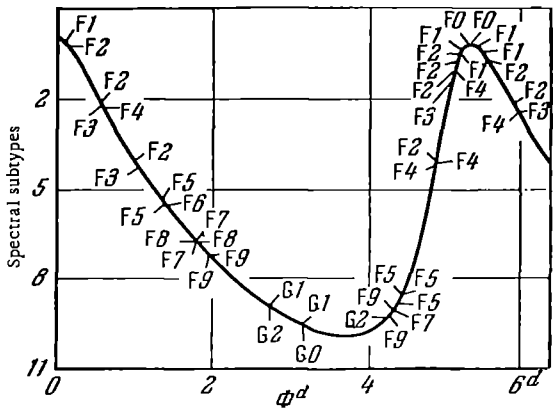


FIGURE 42

Adams and Joy obtained in 1918 at the Mount Wilson Observatory high-dispersion spectra of a number of cepheids near the maximum and the minimum. Having determined the spectral type, they reached the unexpected conclusion that only the spectrum derived from the hydrogen lines showed any significant variation. The spectrum obtained from the metal lines remained largely unchanged. However, Mel'nikov showed in 1944 that this result of Adams and Joy stems from a misunderstanding. A rapid change in the intensities of the metallic lines is observed slightly before the minimum, and not at the minimum proper. Therefore, the spectra taken at the very minimum generally do not reveal any rapid change in the metallic line intensities and thus create a false impression of constant spectrum. Figure 43 shows the spectrum variation curves of δ Cep and η Aql obtained from hydrogen (Balmer, continuous curves) and metal (Ca I, Fe I, Sr II, dashed curves) lines (Mel'nikov's data). Table 3 is a comparison between the findings of Adams and Joy and Mel'nikov.

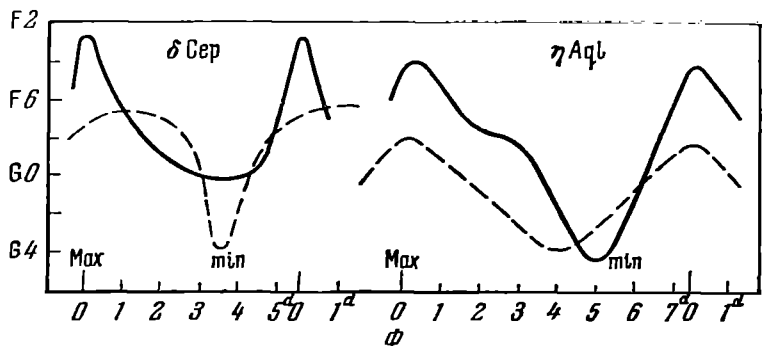


FIGURE 43

TABLE 3

Star	Author; phase and characteristic (HD)							
	Adams and Joy				Mel'nikov			
	maximum		minimum		maximum		minimum	
	hydrogen	metals	hydrogen	metals	hydrogen	metals	hydrogen	metals
δ Cep	F2	F9	F9	G0	F2.4	F6.9	G0.0	G0.9
η Aql	F3	F9	F8	G1	F2.4	F7.8	G4.0	G2.5

We see from Table 3 that if the spectral type is determined near the maximum and the minimum only, then indeed only the hydrogen spectrum shows variation. If, however, the variation of the spectrum is followed for all the phases, the metallic spectrum is found to be equally variable. In this case, however, a later-type average spectrum is obtained, not unlike the situation observed for A stars with metallic lines. As we know, in terms of metallic lines these stars are generally classified as F type. The average data of Figure 43 are listed in Table 4.

TABLE 4

δ Cep		η Aql	
Phase	Spectrum (HD)	Phase	Spectrum (HD)
$0^d.0$	F5	$0^d.0$	F6
0 .5	F6	0 .5	F6
1 .0	F7	1 .0	F7
1 .5	F8	1 .5	F8
2 .0	F8	2 .0	F8
2 .5	F9	2 .5	F9
3 .0	G1	3 .0	G0
3 .5	F3	3 .5	G1
4 .0	F0	4 .0	G2
4 .5	F9	4 .5	G3
5 .0	F7	5 .0	G3
		5 .5	G2
		6 .0	F0
		6 .5	F8
		7 .0	F6

We see from the table that the spectra of δ Cep and η Aql vary from F5 to G3 (hydrogen) and from F6 to G3 (metals).

Among the recent results on the spectral types of long-period cepheids we should mention those of C. Payne-Gaposchkin (1951), who used a uniform sample (slit spectrograms of the Mount Wilson Observatory), the results of O. Struve and Code (1947) from Yerkes slit spectrograms, and of Kukarkin (1944) from cumulative data. Code's determination follows the MKK spectral system (Morgan, Keenan, Kellman). His results for 18 stars are listed in Table 5.

The spectra of cepheids were found to contain stronger hydrogen lines (especially near the maximum) than the spectra of ordinary stars of the same spectral types. The CN band near 4215 Å is conversely weaker

(O. Struve) at the minimum. The spectra of cepheids are highly reminiscent of the spectra of supergiants. This explains the previous name of pseudo-cepheids for the supergiants. These stars include, e.g., α Per (F5Ib), γ Cyg (F8Ib), α Aqr (G2Ib), and others. In the MKK system, the long-period cepheids have spectral types from F to K and luminosity classes from Ia to II.

TABLE 5

Star	lg P	Ampli-tude	Sp (M) MKK	Sp (m) MKK	ΔSp
SU Cas	0,29	0m.38	F5 I—II	F7 I—II	0.2
DT Cyg	0,40	0 .39	F5,5 I—II	F7 I—II	0.15
SZ Tau	0,50	0 .60	F6 Ib	F9 Ib	0.3
V386 Cyg	0,72	1 .00	F5 Ib	G1 Ib	0.6
δ Cep	0,73	0 .72	F5 Ib	G2 Ib	0.7
MW Cyg	0,78	1 .00	F8 Ib	G1 Ib	0.3
η Aql	0,86	0 .71	F6 Ib	G4 Ib	0.8
VY Cyg	0,90	1 .20	F6 Ib	G1 Ib	0.5
S Sgr	0,92	1 .02	F6 Ib	G5 Ib	0.9
BZ Cyg	1,01	0 .80	F8 Ib	G5 Ib	0.7
ζ Gem	1,01	0 .37	F7 Ib	G3 Ib	0.6
Z Lac	1,03	1 .36	F6 Ib	G6 Ib	1.0
TX Cyg	1,17	1 .90	F5 Ib	G6 Ib	1.1
SZ Cyg	1,18	1 .30	F8 Ib	G8 Ib	1.0
X Cyg	1,21	1 .56	F7 Ib	G8 Ib	1.1
CD Cyg	1,23	1 .50	F8 Ib	K0 Ib	1.2
T Mon	1,43	0 .95	F7 Ia—Ib	K1 Ia—Ib	1.4
SV Vul	1,63	0 .90	F7 Ia	(K0)	1.3

The average spectral type at the maximum on the whole varies progressively from F4 to F8, and the spectrum at the minimum shows a much more pronounced variation from F7 to K0 (on passing from long-period cepheids of short periods to those of really long periods). These results are plotted in Figure 44 from the data of Payne-Gaposchkin (1951).

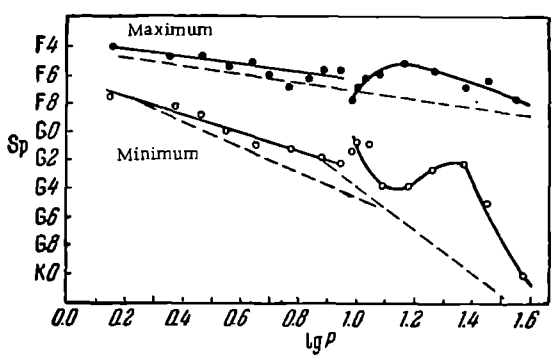


FIGURE 44

More recent studies of the period—spectrum relation include those of Lebedinskii and Gurevich (1955) and, in more detail, P. P. Parenago (1955).

Parenago combined scattered data on 52 long-period cepheids into a uniform system and derived the average period—spectrum relation (see Table 6).

TABLE 6

lg P	Sp (M) HD	Sp (m) HD	$\overline{Sp}$	Number of stars
0.43	F5.5	F8.2	F6.8	5
0.59	F5.2	F9.2	F7.2	5
0.68	F5.2	G1.6	F8.4	5
0.73	F5.2	G1.8	F8.5	5
0.79	F6.3	G2.7	F9.5	6
0.87	F6.3	G2.3	F9.3	6
0.99	F6.8	G2.7	F9 8	5
1.11	F6.0	G6.6	G1.3	5
1.23	F7.4	G9.4	G3.4	5
1.46	F7.4	K1.0	G4.2	5

The data of this table are marked by dashed lines in Figure 44. There is clearly a satisfactory qualitative fit between the two independent determinations. The data presented by solid curves are based on a more uniform material.

The spectrum variation of cepheids is further corroborated by the variation of their color. This points to a high relative significance of the temperature amplitude $A(\lambda,T)$ in the observed light-curve of long-period cepheids. The contribution from the amplitude A_R associated with surface pulsation is relatively small.

Note that the determination of the spectral type of cepheids, like that of normal stars, is highly sensitive to the instrument used for this purpose. The application of a particular instrument often leads to significant deviations in the results of individual observations. The situation is so serious that we should actually speak of peculiar "systems" of spectral types for each instrument. From this point of view, the spectral types and luminosity classes of Table 5, determined in the MKK system (with similar or identical instruments), seem to be the most "objective."

Since the color of long-period cepheids is variable, their spectrum variation curves can be successfully replaced by light variation curves. Both these characteristics are related to a certain extent to the variation of the effective temperature of the star. Numerous studies of the color — period and spectrum — period relation have been published. The most recent results include those of Stibbs (1955), who used the published observations of P, V colors and MKK spectral types to derive for the normal, i.e., not distorted by interstellar and atmospheric absorption, colors of groups A, B, and C at the maximum $(P-V) = -0.40+0.12\mathrm{Sp}\,(M)$ and for groups A and B at the minimum $(P-V) = -0.11+0.063\mathrm{Sp}\,(m)$. The corresponding spectrum variation, starting with type F0, was found to be

$$\mathrm{Sp}\,(M) = 4.8 + 1.5\lg P,$$
$$\mathrm{Sp}\,(m) = 3.0 + 12.3\lg P.$$

Hence it follows that the spectral type at the maximum changes insignificantly (see Figure 44). This led to the conviction that the period—color relation should be considered separately for the maximum and the minimum, and not on the average as before. More recent studies of the normal color—period relation include those of Mel'nikov (1948), Badalyan (1949), Fedorovich (1950), Eggen (1951), Parenago (1955), Stibbs (1955), and others.

Mel'nikov obtained for the median normal colors in the International System

$$CI_0 = 0^m.27 + 0^m.29\lg P.$$

Stibbs introduced painstaking corrections for galactic absorption and obtained from Eggen's observations in the $P-V$ system, which is close to the International System, the following expression for the mean color:*

$$CI_0 = (P-V)_0 = +0^m.13 + 0^m.43 \lg P.$$

Eggen's original colors markedly deviate from this dependence and can hardly be used as a characteristic of the cepheid color. This problem, however, requires further study. One of Badalyan's recent papers (1946) is devoted to this subject.

In 1958 Zaitseva obtained the "hormal color" (corrected for distorting effects) in the Morgan and Johnson $B-V$ system from Eggen's data $(B-V)_0 = 0^m.087 + 0^m.377 \lg P$.

A substantial difficulty in this case is the correction for selective galactic absorption. The recent study of Canavaggia (1955) proved highly promising in this respect: the two-color diagram was used to separate between the contribution from interstellar reddening (the constant component of the cepheid color) and the contribution from temperature variation (the variable component). This method was first applied to normal stars whose color is also inevitably distorted by interstellar absorption. Canavaggia plotted the $G-V$ color (5700—4220 Å) along the vertical and the $G-I$ color (5700—10,300 Å) along the horizontal axis, using Stebbins's six-color photometry of δ Cep and η Aql (and also of some supergiants). The true normal colors obtained in this way provided a check on the variation of radii of these stars, as determined by the previously outlined method. Note that important results can be obtained from photoelectric measurements of $(P-V)$ colors of long-period cepheids in Magellanic Clouds carried out by Gascoygne and Kron (1953). The long-period cepheids in the Small Magellanic Clouds (Population II, as distinct from Population I long-period cepheids in the Large Magellanic Cloud) have the same color at the maximum as the galactic cepheids of groups A, B, and C. This result is consistent with Eggen's remark that long-period cepheids of any type have virtually the same normal color and spectrum at the maximum.

Besides the overall characteristics of the spectra of long-period cepheids discussed above, they also show some fine features. Thus, in 1927 Walton and Shapley reported that Cannon had discovered bright hydrogen lines in the spectra of long-period cepheids of long periods. However, most of these stars were Population II cepheids or W Vir stars (see below). In 1944, Joy and Wilson also found bright lines on the violet side of the H and K Ca II lines in the spectrum of η Aql. Adams (1940) detected fine structure (splitting) of the H and K absorption lines in the spectrum of η Aql.

The study of the spectra and colors of long-period cepheids of known luminosities (see below) make it possible to pinpoint their position on the Hertzsprung—Russell diagram.

3. Radial velocity curves. The determination of the radial velocities of long-period cepheids was begun in the 1890's by A. A. Belopol'skii at Pulkovo. Subsequently further determinations were carried out at various observatories: Lick, Ottawa, Yerkes, Mount Wilson,

* Recently, six-color photometry was used in high-precision measurements of the color of 25 cepheids (Kron and Svolopoulos, 1959). The $(P-V)$ colors are plotted vs. lg P in Figure 84.

Ann Arbor, Bonn, and others. The best curves, however, are the weighted average ones. The observations on the whole reveal fairly good coincidence in the phases of maximum negative and positive velocities (allowing for the star's motion) with light maxima and minima, respectively. The radial velocity curve is thus a sort of mirror image of the light curve. This is clearly visible from Figure 45, which shows the corresponding curves for η Aql. The light variation and radial velocity variation periods are generally equal. Integration over the radial velocity curve obviously gives the change in the star's radius ΔR reckoned from some mean value, say. Differentiation of the radial velocity curve gives the variation of acceleration on the star's surface. Both differentiation and integration are generally carried out numerically. It should be noted, however, that the resulting velocity is an average figure integrated over a sphere, and definitely not the velocity of expansion or contraction of the star. Some uncertainty creeps into the results on account of the unknown limb darkening of the star. A study of the line profiles in expansion and contraction may shed some light on this problem. This, however, requires high-resolution and high-disperison spectrograms.

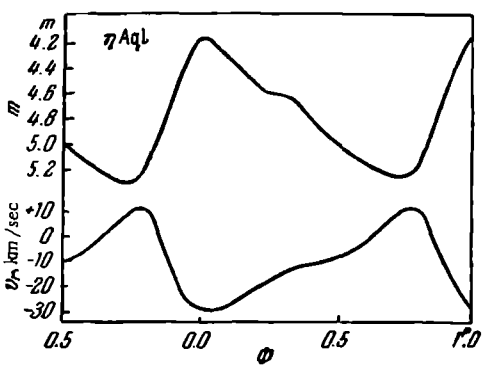


FIGURE 45

Let V_R be the radial velocity of a star at a given point of the disk, and $v = \frac{dR}{dt}$ the true instantaneous velocity of expansion or contraction (i.e., pulsation) of the atmospheric layers of the star. Let further θ be the angle between the line of sight and the radius of the star at a given point on its surface. Clearly,

$$V_R = \frac{dR}{dt} \cos\theta = v \cos\theta.$$

Integration over the disk gives the average velocity

$$\overline{V}_R = \overline{\frac{dR}{dt} \cos\theta} = 2\pi R^2 \int_0^{\frac{\pi}{2}} \frac{v \cos^2\theta \sin\theta \, d\theta}{\pi R^2} = \frac{2}{3} v = \frac{2}{3} \frac{dR}{dt}.$$

Hence,

$$v=\frac{dR}{dt}=\frac{3}{2}\overline{V}_R.$$

However this result is valid only for a disk with uniform brightness at all points.

If the limb darkening is given by

$$I(\theta)=I\left(\frac{\pi}{2}\right)(1+\beta\cos\theta)=I(0)\left(\frac{1+\beta\cos\theta}{1+\beta}\right)=I(0)(1-u+u\cos\theta),$$
$$u=\frac{\beta}{1+\beta},$$

we have

$$\overline{V}_R=\overline{v\cos\theta}=v\frac{\int_0^{\frac{\pi}{2}}(1+\beta\cos\theta)\cos^2\theta\sin\theta\,d\theta}{\int_0^{\frac{\pi}{2}}(1+\beta\cos\theta)\cos\theta\sin\theta\,d\theta}=v\frac{4+3\beta}{6+4\beta}.$$

For $\beta = 3/2$ ($u = 3/5$), which corresponds to a "gray body," we get

$$\overline{V}_R=\frac{17}{24}v=\frac{17}{24}\cdot\frac{dR}{dt}.$$

Hence

$$v=\frac{dR}{dt}=\frac{24}{17}\overline{V}_R.$$

For a gray body with $u = 3/5$ ($\beta = 3/2$) we thus have*

$$\frac{dR}{dt}=v=\frac{24}{17}(v_R-v_\gamma)=\frac{24}{17}\overline{V}_R.$$

Here v_R is the radial velocity obtained from observations, v_γ is the velocity of the center of gravity of the star, equal to $\frac{1}{P}\int_0^P v_R\,dt$. Hence

$$R_P=R_m+\int_{t_m}^{t_P}v\,dt=\text{const}+\int_{t_m}^{t_P}v\,dt=\text{const}+\int_{t_m}^{t_P}\frac{dR}{dt}\,dt.$$

The integration constant is so chosen that the time-average change in the radius $\Delta R=R-\overline{R}$ is zero; in other words $\overline{R}$ is defined as the average radius. The data for η Aql are listed in Table 7 (M. Schwarzschild, B. Schwarzschild, and W. Adams, 1948).

The calculated values of $\frac{dR}{dt}$ and $\Delta R=R-\overline{R}$ for δ Cep and η Aql are plotted in Figure 46 (Stebbins, 1953).

We see from the figure that the minimum and maximum radius do not coincide with the maximum and minimum magnitude, at variance with the original results of the pulsation theory in its initial form. The observations indicate that the minimum radius (maximum compression) precedes the

* For $u = 0$ and 1 (i.e., $\beta = 0$ and ∞), the coefficients are respectively 24/16 and 24/18.

maximum magnitude by $P/4$ or by $\frac{\pi}{2}$. This widely known phase shift was generally attributed to a peculiar behavior of the cepheid atmospheres, immersed in the variable radiation field of the photosphere. Indeed, if the star's radius is calculated only from the color and the temperature (strictly speaking, the effective temperature T_e), using the equality $m = -2.5 \lg R^2T^4 + \text{const}$, the above phase shift is virtually eliminated. This is understandable, since in this case we calculate the variation of the photospheric radius, i.e., the radius of the layers whose pulsation is actually observed in light measurements.

TABLE 7

Phase	Δm_{pg} ηAql−βAql	$\frac{dR}{dt}=\dot{R}$, km/sec	$\Delta R=R-\bar{R}$, 10^{10} cm	$\frac{d^2R}{dt^2}=\ddot{R}$, cm/sec^2	Phase	Δm_{pg} ηAql−βAql	$\frac{dR}{dt}=\dot{R}$, km/sec	$\Delta R=R-\bar{R}$, 10^{10} cm	$\frac{d^2R}{dt^2}=\ddot{R}$, cm/sec^2
0.00	−0.42	+24.9	−16.2	−1	0.60	+0.60	−13.7	+12.5	−24
.04	− .35	+23.7	−10.1	−7	.64	+ .67	−19.8	+8.4	−25
.08	− .26	+21.6	−4.5	−10	.68	+ .70	−26.1	+2.7	−26
.12	− .16	+18.9	+0.5	−11	.72	+ .67	−31.8	−4.5	−17
.16	− .08	+16.0	+4.9	−12	.76	+ .58	−34.0	−12.8	+2
0.20	−0.01	+13.0	+8.4	−11	0.78	+0.50	−33.2	−17.0	+14
.24	+ .01	+10.2	+11.3	−11	.80	+ .42	−30.3	−20.9	+32
.28	+ .02	+7.5	+13.5	−10	.82	+ .32	−25.2	−24.4	+49
.32	+ .04	+5.1	+15.1	−9	.84	+ .22	−18.2	−27.1	+65
.36	+ .13	+3.0	+16.0	−7	.86	+ .11	−9.4	−28.8	+74
0.40	+0.28	+1.4	+16.6	−5	0.88	0.00	0.0	−29.4	+76
.44	+ .37	+0.6	+16.8	−2	.90	− .10	+8.9	−28.8	+64
.48	+ .43	−0.3	+16.9	−7	.92	− .22	+15.7	−27.3	+47
.52	+ .48	−3.1	+16.5	−16	.94	− .32	+20.5	−25.0	+31
.56	+ .54	−7.9	+15.2	−22	0.98	−0.44	+24.7	−19.3	+5

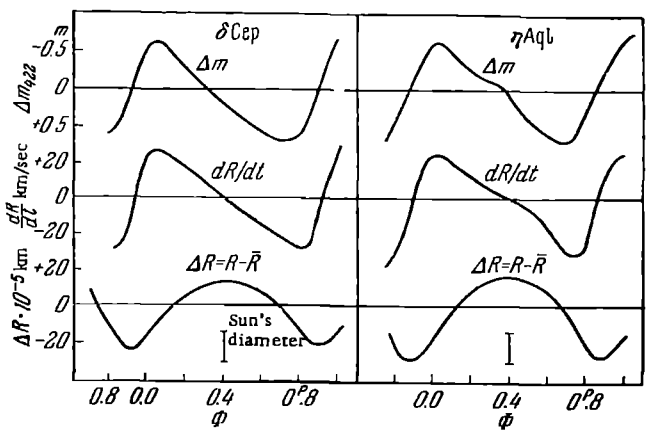


FIGURE 46

Note that the very integration of the radial velocity curve may give rise to serious doubts. After all, in observing the radial velocity of a star, we do not measure the motion of the same atoms but only the displacement of the "active surface" where the lines effectively originate.

If the Stefan—Boltzmann law is assumed (which is apparently justified, if we are to judge from data on eclipsing binaries with known radii, luminosities, and temperatures), the variation of the bolometric magnitude of a cepheid can be written in the form*

$$A_b = \Delta m_b = -2.5 \lg \frac{R_m^2 \cdot T_m^4}{R_M^2 \cdot T_M^4} = 4.34 \frac{\Delta R}{\overline{R}} - 8.70 \frac{\Delta T}{\overline{T}}$$

Here $\delta R = 2\Delta R$ and $\delta T = 2\Delta T$ are the total amplitudes of the radius and the temperature (ΔR and ΔT being the respective half-amplitudes). Hence,

$$\Delta m_b = A_b = 2.17 \frac{\delta R}{\overline{R}} + 4.35 \frac{\delta T}{\overline{T}} = \Delta m_R + \Delta m_T.$$

ΔR can be found by integration of the radial velocity curve (see Figure 48). Given this quantity, we can calculate the temperature component of light variation (its amplitude being Δm_T) assuming the surface to remain invariable.

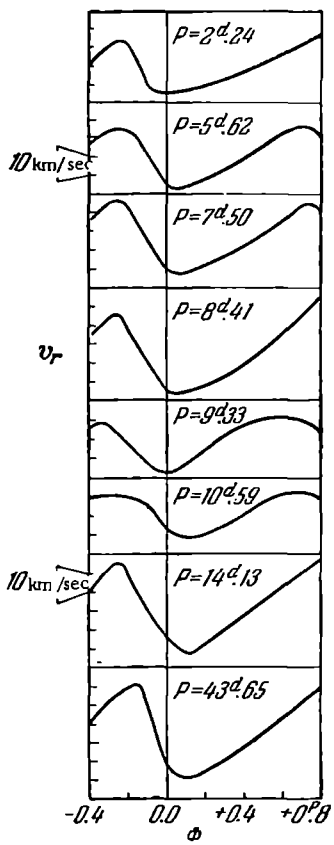


FIGURE 47

Figure 47 gives the average radial velocity curves of long-period cepheids of various periods (Payne-Gaposchkin, from material obtained at the Mount Wilson Observatory, 1951).

We see from the figure that the radial velocity curve is not always a precise mirror image of the light curve.

The results point to a certain dependence of the phase of the maximum negative velocity on the period. Similar phase relations have been known for some time now. More recent data were obtained by Payne-Gaposchkin in 1951.

For comparison, Figure 48 gives Payne-Gaposchkin's data on the relation of the amplitudes of light, radial velocity, and spectrum variation to the period. The figures reveal a certain fit between the maximum and minimum phases, as well as a progressive variation, reminiscent of the variation of the amplitudes with the period. The figures also show that periods near 10 days ($\lg P \approx 1.0$) are critical for a whole range of cepheid characteristics. Indeed, we have shown in the preceding that deviations from other fundamental characteristics are also observed near this period: the light curves become skewed, significant fluctuation of the period is observed, the period—luminosity relation breaks down (see below), and so on.

Note in particular that the period—light variation amplitude relation, the period—color (or spectrum) variation amplitude relation, and other correlations follow a much more regular (almost linear) trend if only photoelectric

* Series expanding the functions $\frac{1}{1+x}$ and $\lg(1+x)$ for $x^2 < 1$ and $|x| < 1$, respectively ($R_m = \overline{R} + \Delta R$, $R_M = \overline{R} - \Delta R$; $T_m = \overline{T} - \Delta T$, $T_M = \overline{T} + \Delta T$).

observations are used. Moreover, various relations in cepheids remain fairly inconspicuous if lg P is used as the argument, while conversely they emerge as a strict functional dependence if the exact magnitude range is used as the argument.

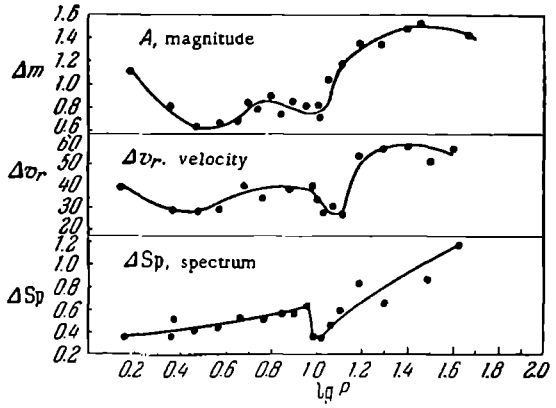


FIGURE 48

4. Surface acceleration curves. Back in 1934 Kipper showed that the surface acceleration of a pulsating star — a cepheid — can be calculated from the equality

$$g = g_{ap} = g_0\left(1 + 2\frac{\Delta R}{R}\right) - \frac{24}{17}\frac{dv_R}{dt} = g_0\left(1 + 2\frac{\Delta R}{R}\right) - \frac{d^2R}{dt^2}.$$

Here the first term is the surface gravitational acceleration $g_0 = G\frac{\mathfrak{M}}{R^2}$ and its variation due to the change in the star's radius. The second term incorporates the additional acceleration associated with radial pulsations of velocity $v = \frac{dR}{dt}$ (the observed radial velocity being v_R). This additional acceleration can be found, as we have noted before, by numerical differentiation of the radial velocity curve (Table 7). The variation of $\frac{24}{17}\frac{dv_R}{dt} = \frac{d^2R}{dt^2}$ for η Aql, derived from the radial velocity curve (using a combination of Lick and Mount Wilson data) according to Table 7, is shown in Figure 49.

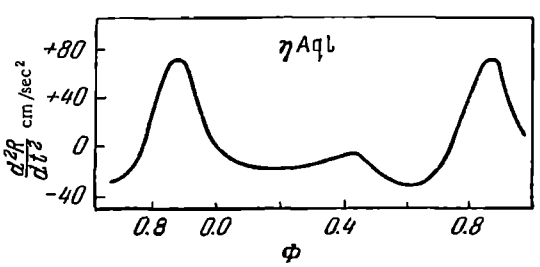


FIGURE 49

Further information can be obtained from line intensities in cepheid spectra (neutral and ionized elements). Indeed, having determined the degree of ionization in the atmospheres of δ Cep and η Aql, say, from the Ca I and Ca II lines, we can find the variation of electron pressure at fixed temperature (strictly speaking, the ionization temperature). Now, adopting a certain model of the atmosphere, we obtain a relation between electron pressure and surface acceleration and can thus find the variation of surface acceleration.

On the other hand, using the preceding equation we obtain the variation of the relation acceleration compared to that at the light maximum, g_M:

$$\frac{g}{g_M} = \left(\frac{R_M}{R}\right)^2 - \frac{1.5\,\frac{dv_R}{dt}}{g_M}.$$

The factor $1.5\frac{dv_R}{dt}$ is obtained from the radial velocity curve. Taking from the theory $\frac{\Delta R}{R} \approx 0.1$ and using some g_M, we plot a family of curves

$$\frac{g}{g_M} = f[\Phi]$$

(where Φ is the phase).

The curve showing the best fit with observations (using the degree of ionization) gives the true value of g_M. The results for δ Cep and η Aql are shown in Figure 50. The numerical values of the surface acceleration for δ Cep and η Aql were found to be 20 and 17 cm/sec^2, respectively. The dynamic value $(g_M)_d = G\frac{\mathfrak{M}}{R^2}$ was approximately found to be 300 and 200 cm/sec^2 for δ Cep and η Aql (from the mass—luminosity relation for normal stars with Mel'nikov's zero point $\Delta M_0 = -0^m.5$). This result is not unexpected. We know that for normal supergiants the acceleration calculated from spectral lines is invariably less than the dynamic acceleration.

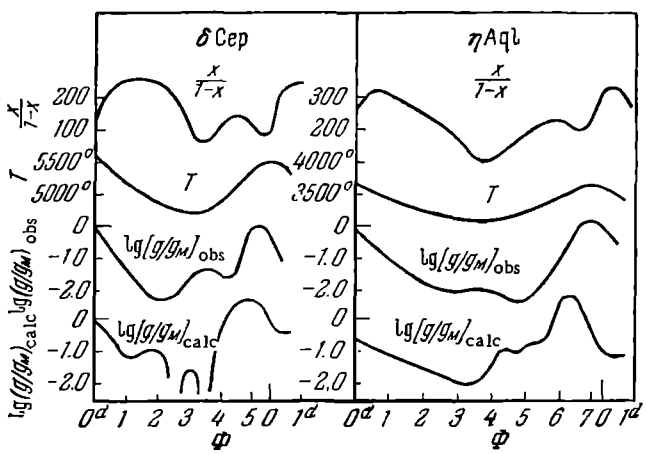


FIGURE 50

Mel'nikov (1954) showed for A stars that radiation makes a substantial contribution to the overall effect associated with the large extent of the atmospheres (which lower the surface acceleration, $g = G\frac{\mathfrak{M}}{(R+H)^2}$, H being the thickness of the atmosphere).

Thus, the g_M obtained from spectral lines provides only a lower-bound value for the dynamic g_d. Moreover, note that the applicability of the mass—luminosity relation to long-period and short-period cepheids has not been conclusively justified by observations to this day. Spectrophotometric data are also somewhat ambiguous. For example, Pannekoek (1946) estimated that g was 1/10 of g_d. M. Schwarzschild (1948), on the other hand, studied this problem for the particular case of η Aql and obtained from spectrophotometric data $g = 32\,\mathrm{cm/sec^2}$ for the line profiles determined by large-scale turbulence and $g = 8\,\mathrm{cm/sec^2}$ for line profiles determined by the star's spin. At present, we cannot distinguish between these two cases from spectrophotometric data. From the mass—luminosity relation, taking $\Delta M = -1^m.5$, we get $g_d = 55\,\mathrm{cm/sec^2}$ for η Aql (Savedov, 1953). Thus, the differences in g and g_d can be gradually eliminated or reduced for cepheids.

The radial velocity curves of long-period cepheids can be represented by two straight lines for the ascending and the descending branch. Using this representation, Mel'nikov (1948) obtained for the masses of long-periodic cepheids

$$2{,}1 < \frac{\mathfrak{M}}{\mathfrak{M}_\odot} < 8.0.$$

In 1955 Parenago obtained by an entirely different method for the masses of long-period cepheids $\frac{\mathfrak{M}}{\mathfrak{M}_\odot} = 6$ on the average.

Proceeding from this result and other arguments, Parenago accepted Mel'nikov's hypothesis of constant masses of long-period cepheids. At this point, we should recall an important result obtained by Thiessen in 1956. He studied the remarkable binary BM Cas which comprises an A5 supergiant (one of the stars of highest luminosity in the sky $M = 8^m.4$) and a long-period cepheid with $P = 27^d$ ($M = -6^m.0$). In this exceptional case the mass of the cepheid could be determined directly. It was found to be $\frac{\mathfrak{M}}{\mathfrak{M}_\odot} = 14.3$. The mass—luminosity relation gives for this rate 23 or even 29 (if we use Petrie's result, 1950). The mass—luminosity relation thus seems to be indeed inapplicable to cepheids, since their true masses are markedly less.

Note that the differential shift of lines in the spectra of cepheids has been studied by numerous authors. Adams obtained a dependence of this shift on the atomic excitation potentials. The other authors were mainly looking for the dependence of the differential line shift on the depth of line formation in the atmosphere. The depths were taken from the data of solar eclipses, but since the Sun is a dwarf star the applicability of these data is not entirely justified. The shifts themselves are fairly small. Better results are apparently those recently published by Jacobsen in revised form. However, high-dispersion spectrograms of η Aql show (M. Schwarzschild and others, 1948) that in many cases the differential

shift is attributable to blends. Measurements of the shift of Fe I and Fe II lines gave equal differential shift of -0.4 ± 0.6 km/sec. Yet it is entirely obvious that the lines of neutral and ionized atoms of the same element, having different excitation potentials, effectively originate at different depths in the atmosphere. In the author's opinion, this is associated with the fact that the atmospheric density of the long-period cepheids studied is sufficiently high compared to the lower (critical) density required to produce mutual drag between moving atoms. Note that solar prominences show virtually no differential shift.

Kukarkin carefully re-examined this problem and showed in particular (1936) that in the atmosphere of δ Cep the radial velocities measured from spectral lines originating at smaller depths in the atmosphere show a certain phase lag. The amplitude of radial velocity variation shows a certain tendency (though within the margin of error) to decrease in higher lying layers of the atmosphere. On the whole, the shape of the radial velocity curve obtained from different lines is somewhat different.

Another point to remember is the phase difference between the maximum negative radial velocity (derived from the average curve based on all the spectral lines) and the phase of the light maximum. This phase shift is furthermore correlated with the period in a highly complex fashion.

In 1949 Jacobsen, like Adams before him in 1940, discovered bright violet satellites of H and K Ca II lines in the spectrum of η Aql. Immediately after that, he carried out new measurements of the differential shift of individual lines in this spectrum with dispersion of 20 Å/mm (Victoria, 1950). Using these emission (bright) lines, the core of the absorption K Ca II line, and six other metal lines in the spectrum of η Aql, Jacobsen determined the velocities of the star relative to the Sun, which were found to be −47, −46, +22, −28, −34 and −4 km/sec. In the spectrograms these lines had equivalent widths of 0.27, 0.51, 0.28, 10.1, 10.3 and 0.1 Å, respectively. Somewhat later, namely in 1952, Grandjean and Ledoux reported differential line shifts in the spectra of η Aql (near the phases $0^P.013$ and $0^P.556$) and ζ Gem (near $0^P.017$ and $0^P.55$) from the data of Mount Wilson Observatory. These shifts, according to the authors, were 1.5—2.0 km/sec.

The "fine structure" of the radial velocity curves of long-period cepheids clearly have a noticeable effect on the surface acceleration g, which is calculated from these curves, and the results of these calculations must be considered with some care. The effective acceleration can be found also from the absorption line intensities, in particular if we associate the broadening of these lines with the Stark effect.

5. Line intensity curves. Visual estimates of line intensities in cepheid spectra using slit spectrograms, begun by Lehman and Belopol'skii, were also carried out by Ogorodnikov (δ Cep, 1933), Pariiskii (η Aql, 1935), Hughes (various cepheids, 1931), and many others. More exact spectrophotometric measurements of line intensities in the spectra of long-period cepheids were carried out by Krüger (1937), Mel'nikov (1939, 1944), Walraven (1948), M. Schwarzschild, B. Schwarzschild, and W. Adams (1948), and many others.

In 1944 Mel'nikov established that the lines in the spectra of δ Cep and η Aql were highly asymmetric with a large violet wing. This skewness leads to biased errors in radial velocities, especially in visual measurements with a micrometric microscope. The errors prove to be negative:

they amount to a few km/sec. This result was also confirmed by Mel'nikov in a statistical study of the radial velocities of long-period cepheids. His findings give a negative K term reaching on the average − 5 km/sec.

These results point to a possible ejection of atoms from the extended atmospheres of the long-period cepheids with markedly reduced surface accelerations. If this is indeed so, we have a satisfactory explanation why the active atmospheric turbulence in these cepheids (see below) fails to heat up the atmosphere: the atoms escaping from the atmosphere carry off part of the excess energy. The ejection of gas for δ Cep is corroborated by the fact that its radial velocity is a few km/sec less than the velocity of its physical dwarf companion (a B8—B9 star). The velocities, unfortunately only from W. Luyten's data, are −16.4 and −11.0 km/sec for δ Cep A and δ Cep B, respectively (the orbital velocity from statistical considerations is at most 2 km/sec for a separation of 40″). The companion δ Cep B is a $6^m.01$ star with color $C_p = -0^m.12$ in Eggen's system.

Back in 1944, Adams discovered a complex line structure in the spectrum of η Aql (presence of separate components). This structure, resolved in Coudé spectrograms, is at least partially responsible for the observed skewnéss of the lines. Mount Wilson Observatory data on η Aql were studied in 1952 by van Hoof and Deurinck, who found that the skewness of the lines varied in agreement with the pulsation theory and even permitted establishing the limb darkening of the disk of η Aql. The darkening coefficient was found to be on the average $u \approx 0.6$ or $\beta \approx 3/2$, i.e., as for a gray body. Moreover they found that the ratio of the pulsation velocity $v = \frac{dR}{dt}$ to the radial velocity v_R was a function of line intensity, so that the conversion factor from $\bar{V}_R$ to v was variable. Similar fine features provided a logical explanation of the difference in the radial velocities, or alternatively in the differential shifts of the F I and F II lines.

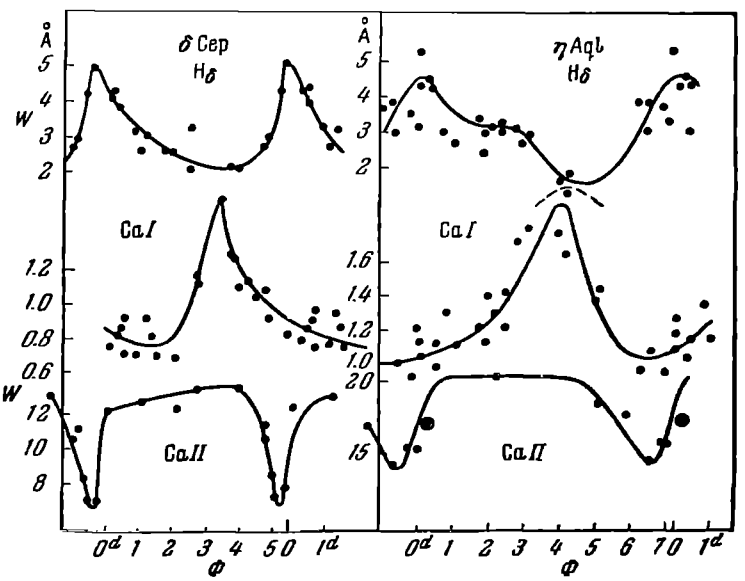


FIGURE 51

Mel'nikov used in 1944 the δ Cep and η Aql spectrograms obtained at Pulkovo (10 Å/mm dispersion) and Simeise (36 Å/mm). Figure 51 plots the variation of the equivalent widths of some lines with the phase of light variation. W. Adams, M. Schwarzschild, and B. Schwarzschild studied (1948) the spectrograms obtained with a more powerful Coudé spectrograph (2.9 Å/mm dispersion). They mainly measured the line "depths", defined as $\lg D = \lg \frac{I_0}{I_c} = \lg \frac{1}{r_0}$ where I_0 is the continuum intensity, I_c is the line center intensity, and r_0 is the relative central intensity. The equivalent widths were actually obtained only for selected lines and phases. A similar spectrophotometric study of δ Cep from the data of Pannekoek was carried out by Walraven in 1948 (a three-prism spectrograph, Victoria Observatory). Mel'nikov plotted the curve of growth for the mean phase of δ Cep and η Aql from numerous lines of a whole range of elements. This curve gave the abundance, i.e., the relative content, of the chemical elements in the atmospheres of δ Cep and η Aql. The total content of various atoms in the atmospheres of δ Cep and η Aql in a column of 1 cm^2 cross section is listed in Table 8. For comparison Table 9 gives the relative content of the chemical elements in the atmospheres of cepheids, one supergiant, and an irregular variable.

TABLE 8

Element	Atomic number	δ Cep	η Aql	Element	Atomic number	δ Cep	η Aql
		lg *N*	lg *N*			lg *N*	lg *N*
H	1	25.5	30.30	V	23	16.92	18.65
C	6	(20.85)	(24.27)	Cr	24	16.95	18.44
Na	11	19.39	20.97	Mn	25	16.65	18.13
Mg	12	18.82	20.10	Fe	26	18.56	20.02
Al	13	15.90	16.87	Ni	28	18.17	20.49
Ca	20	17.87	18.41	Sr	38	14.80	15.53
Sc	21	14.09	14 43	Zr	40	14.82	16.09
Ti	22	16.98	17.75	Ba	56	13.86	14.51

TABLE 9

Element	Stars and spectra (HD)			
	δ Cep (cF9)	η Aql (cG0)	γ Cyg (cF7)	RCrB (cFp)
H	100%	100%	99%	27%
C	0.001	0.0001	0.5	69
Metals	0.001	0.0000001	0.5	4

We see from the table that the abundance of hydrogen in the atmospheres of the two cepheids δ Cep and η Aql is anomalously large. The content of other elements in a number of objects is closer to the normal. The same growth curves gave by the ordinary techniques the excitation temperatures T_{ex} of the atoms and the electron pressure P_e for the mean light-variation phase. The results gave

(a) for δ Cep

$$T_{ex} = 5270° \pm 370°,$$
$$P_e = 2.4 \text{ bar};$$

(b) for η Aql

$$T_{ex} = 3550° \pm 250°,$$
$$P_e = 1.6 \cdot 10^{-4} \text{ bar}.$$

The P_e for η Aql is obviously much too low.

Confining his construction only to the middle part of the growth curve for all the phases of δ Cep and η Aql, Mel'nikov established the turbulent state of the atmospheres of these stars and determined the eddy velocities at various phases. The results are plotted in Figure 52.

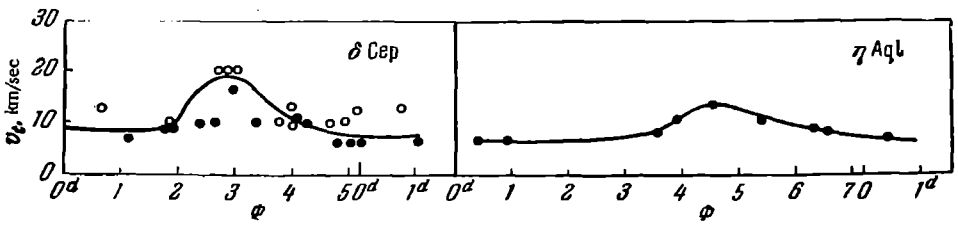


FIGURE 52

The figure actually plots the total velocity $V = \sqrt{V^2_{therm} + V^2_{turb}}$ (V_{therm} being the termal and V^2_{turb} the eddy velocity of motion in the atmosphere), but at the typical cepheid temperatures V_{therm} is of the order of 1 km/sec for the relevant lines (using weighted values), and therefore $V \approx V_{turb}$.

Schwarzschild and others determined the excitation temperatures and electron pressure from the line depths.

For excitation temperatures of η Aql they obtained

$$5040\left(\frac{1}{T_{ex,m}} + \frac{1}{T_{ex,M}}\right) = 0.19.$$

On the other hand, the cG4 spectrum at the minimum gives an effective temperature $T_{e,m}$ = 4800°K (Kuiper's scale of effective temperatures).

Given T_e and M_v, we can use Hertzsprung's relation say,

$$M_v = 2.3\left(\frac{14\,300}{T_{e,m}}\right)^{0.93} - 5 \lg R + 53.71$$

to obtain $R_m = 2.58 \cdot 10^{12}$ cm. At the minimum, the radial velocity curve (phase $0^p.68$) gives $\Delta R = R_m - \bar{R} = 0.03 \cdot 10^{12}$ and therefore $\bar{R} = 2.55 \cdot 10^{12}$ cm.

At the maximum the same curve gives $\Delta R = R_M - \bar{R} = -0.16 \cdot 10^{12}$ and thus $R_M = 2.39 \cdot 10^{12}$. The visual absolute magnitude of η Aql at the maximum (M_v) is $-2^m.8$; hence using the above relation we get $T_{e,M}$ = 5760°K.

From these two values we obtain

$$5040\left(\frac{1}{T_{e,m}} - \frac{1}{T_{e,M}}\right) = 0.17.$$

This figure virtually coincides with the 0.19 obtained for the excitation temperatures from the growth curve. It therefore seems that the $T_{ex,m}$ and $T_{ex,M}$ determined using 0.19 as the value of the difference of the reciprocals

gives an adequate fit to the amplitude of effective temperature fluctuations and probably also its variation with the light phase (the latter is conveniently represented by the value $\Delta\theta = \frac{5040}{T_e}$).

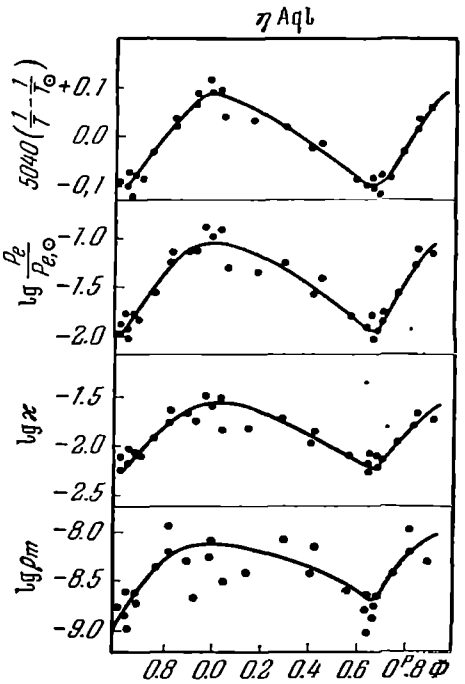


FIGURE 53

To find the absolute effective and hence ionization temperatures, we should determine the absolute value of $T_\odot$. The iron lines in the Sun probably originate at an optical depth of 0.3, where the local temperature is $T_{0.3,\odot} = 5300°$. This gives ionization temperature of 5300° for the middle part of the descending branch of η Aql ($\Delta\theta = 0$). At this phase the effective temperature is also 5300°K. This is the figure actually used. From the first expression for $\lg\frac{X}{X_\odot}$ and the Fe I lines we find α and Δθ.* From the second expression, using these values and the Fe II lines, we find the average δ. Given α, δ, * and Δθ, we can find the relative electron pressure.

These observations also give the continuum absorption coefficient x and the hydrogen to metals abundance ratio (abbreviated as H/M).

The continuous absorption in the atmosphere of η Aql is apparently associated with the negative H^- ions, as is common for G type stars.

For the hydrogen to metals abundance ratio in the atmosphere of η Aql we have

$$H/M = 4000.$$

The available data enable us to calculate the relative variation of the density ρ in the atmosphere of η Aql.

Figure 53 plots the variation of $\Delta\theta, \frac{P_e}{P_{e\odot}}$, lg x and lg ρ_m for η Aql from the data of Schwarzschild et al. (1948). We see from the figure that the maxima of these curves approximately coincide with the maximum of the light curve, whereas the minima somewhat precede the light variation minima.

From the equivalent widths of a number of Fe lines, Schwarzschild et al. plotted the growth curve for η Aql. Using this curve, they determined the velocity of small-scale turbulence in the atmosphere (for the middle part

* As always, $X_0 = 1.5\cdot10^{-2}\frac{NHf}{\Delta\nu_D}$; $\lg\frac{X}{X_\odot} = \alpha + (E_s + 2.43)\Delta\theta$; $\delta = (\chi_i + 0.37)\Delta\theta + \frac{5}{2}\lg\frac{T}{T_\odot} - \lg\frac{P_e}{P_{\odot,e}}$ (E_s, χ_i are the excitation and ionization potentials).

of the descending branch). It was found to be 4 km/sec, which virtually coincides with our value of 6 km/sec (see Figure 81). The thermal velocity, on the other hand, is

$$V_{\text{therm}} = \sqrt{\frac{2kT}{\mathfrak{M}}} = \sqrt{\frac{2kT}{\mu\mathfrak{M}_1}} = \sqrt{\frac{2RT}{\mu}} = \frac{9.3}{\sqrt{\mu}}\,\text{km/sec},$$

where $\mu = \frac{\mathfrak{M}}{\mathfrak{M}_1}$ is the atomic weight* taken equal to 56 for Fe. Thus $V_{\text{therm}} = 1.2$ km/sec. These data, like our data obtained four years previously, point to substantial small-scale turbulence in the atmosphere of η Aql. This small-scale turbulence markedly increases the equivalent line widths. Large-scale turbulence, on the other hand (of the order of the atmospheric layer), does not affect the equivalent line width but distorts the line profile. Schwarzschild et al. obtained velocities of 12.2 km/sec for the large-scale turbulence (corrected for the instrumental profile of 3.5 km/sec) from the average line profile of the weak lines 4440, 4548, 4587, and 4602 Å. Later, in 1952, van Hoof and Deurinck used the Mount Wilson spectrograms to confirm the presence of large-scale turbulence in the atmosphere of η Aql.

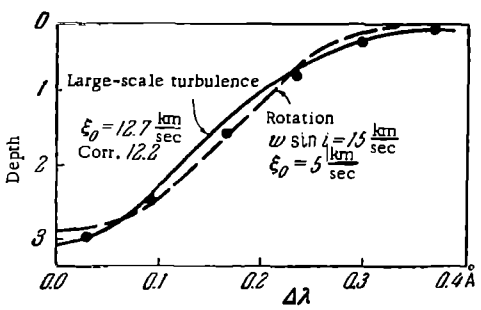


FIGURE 54

The average Fe I line profiles are shown in Figure 54. The Doppler curve fit corrected for the instrumental profile is the solid curve.

Thus the atmosphere of η Aql contains both small-scale and large-scale turbulence. Unfortunately, the large-scale turbulence cannot be isolated from the effect of the radial convection in the atmosphere or the star's spin. The line profile associated with a spin of $w \sin i$ = 15 km/sec (w being the equatorial velocity, i the inclination) is marked by the dashed curve in Figure 54. Note that this average profile is based on four normal profiles, normalized to the measured depth and equivalent width.

The above method of study of atmospheric motions in stars was recently developed in considerable detail by O. Struve and Huang Shu-Su.

§ 43. RR Lyrae stars. There are very few comprehensive studies of the short-period cepheids, commonly known as RR Lyrae stars. Only their light curves have been investigated in some detail. Other characteristics have been poorly studied.

1. Light curves. Figure 55 is one of the latest photoelectric light curves of a typical short-period cepheid — RR Lyrae itself — from Hardy's data (1955). Figure 56 gives some "standard" light curves of RR Lyrae stars of various periods. These curves, like those of the long-period cepheids, are the work of Kukarkin and Parenago. For comparison, the figure also gives two standard light curves of cepheids with over-day

* $\mathfrak{M}_1 = \frac{1}{16}\mathfrak{M}_{O^{16}}$ (i.e., 1/16 of the mass of the most abundant oxygen isotope), $k = \frac{R}{N}$, $\mathfrak{M}_1 = \frac{1}{N}$.

periods. Among the known RR Lyrae stars, the shortest periods are those of SX Phe = HD223065, $P = 0^d.055$ (80 min) and CY Aqr, $P = 0^d.061$ (88 min).

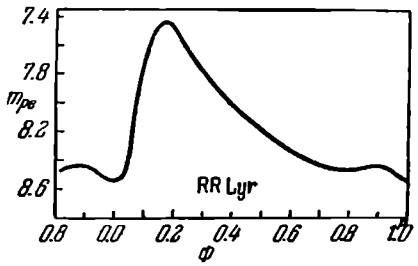


FIGURE 55

We see from Figure 56 that the light curves of RR Lyrae stars are relatively symmetric. According to S. Bailey's classification, these are type *c* curves. Near periods greater than $0^d.4$, the curves become markedly skewed (type *a*). As the period further increases, the symmetry improves (type *b*). The dependence of the shape of the light curve of RR Lyrae stars on the period was discovered in 1902 by Bailey in his study of cluster variables. The short-period cepheids among the cluster variables indeed have skewed light curves for periods of over $0^d.5$ (types *a* and *b* in Bailey's classification), whereas variables with periods less than $0^d.5$ have symmetric curves (type *c*). However, S. Gaposchkin studied RR Lyrae stars in the direction to the galactic center in the region of the cluster NGC 6522 and found variables with periods of some $0^d.4$ (i.e., apparently type *c*) with definitely skewed curves of type *a* and *b*. It thus appeared that the variables observed in the direction to the galactic center were inherently different from the galactic field variables, which did not reveal such anomalies. However, in 1960 D. Alexander made two-color colorimetry of these centerward variables using the 188-cm Radcliffe reflector and found that S. Gaposchkin's periods were insufficiently accurate. If the accurate periods are used, the anomaly is smoothed out and the light curves of RR Lyrae stars in the direction to the galactic center lose their anomalous features, perfectly merging with the variables of the halo component (the galactic field) and the globular clusters. As we have noted, the classification of the RR Lyrae stars into groups *a*, *b*, *c* is to a certain extent analogous to the groups *A*, *B*, *C* of the long-period cepheids. There are, however, essential differences. The light curves of RR Lyrae stars, for example, show no "hump", or secondary wave, which is a characteristic feature of type *B* long period cepheids (η Aql stars).

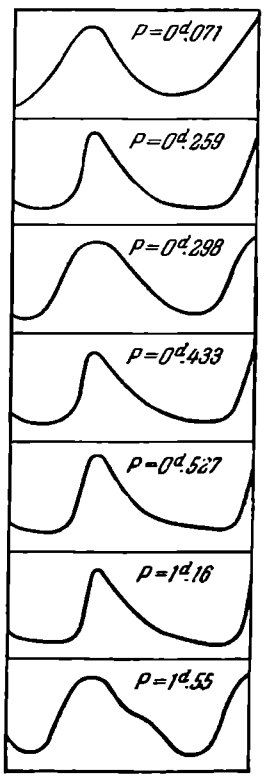


FIGURE 56

Unlike the long-period cepheids, the RR Lyrae stars are characterized by considerable variation of the light curve. AR Her, a variable with a period of $0^d.47$ (11^h17^m), changes its light curve with a period of $31^d.5$ (i.e., periodically every 67 cycles) (Balasz, Detre). A similar effect is observed for RR Lyr, RW Dra, Y Leo, XZ Cyg. The period of light curve variation of

RR Lyr was conventionally taken to be 38^d (67 cycles). However, more exact photoelectric recordings of the light curve obtained by Walraven (1949), who used a highly original technique, gave a revised value $P = 0^d.5668375$ for the fundamental period and values of $P_1 = 72^d.37$ and $P_2 = 41^d.0$ for the variation of the light curve shape and the light maximum. The phase of the maximum fluctuates by $\pm 0^d.02$. For this reason the initial phase was taken as the median phase, and not the phase of the maximum, as the former is less susceptible to fluctuations.

The solid line in Figure 57 gives the variation of the light curve of AR Her during one cycle of $31^d.5$. The dashed curve is the average light curve of the star.

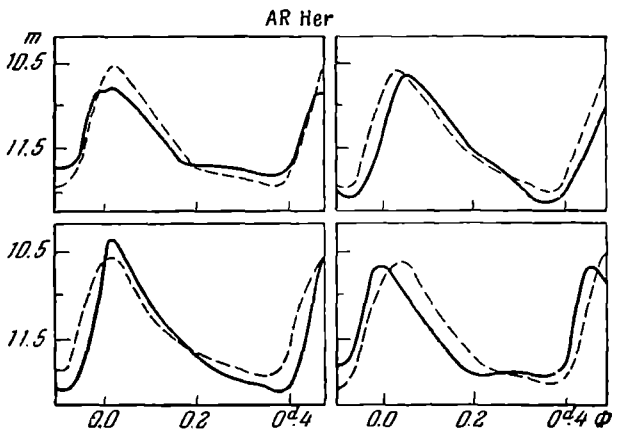


FIGURE 57

Unlike most of the long-period cepheids, RR Lyrae stars are characterized by marked variation of the period. For example, the period of RZ Cep as determined in 1898 was $7^h24^m28^s.76$, but in 1901 a different value, $7^h24^m24^s.78$, was obtained. In 1916 the period was $7^h24^m29^s.11$, and in 1923 $7^h24^m30^s.95$. After that the period remained invariable for a long time. AC And was identified as a binary by Florya (1937). Each of the components of this binary is an RR Lyrae star with periods of $0^d.53$ and $0^d.71$, respectively. The epochs of maximum of RR Lyrae stars are also adequately fitted by parabolic expressions, which points to progressive variation of the periods. However, according to Tsesevich, Detre, and others, this problem is highly uncertain at present.

Most RR Lyrae stars apparently show slow and irregular changes of the period.

It should be further remembered that the results are almost always affected by cumulative errors, as previously noted for the long-period cepheids (T. Sterne, 1934).

CY Aqr and UY Boo, stars of exceedingly short period, have greatly skewed light curves. There is a possibility that this skewness differs for stars in different parts of the Galaxy. It may be more pronounced near

the galactic center. In conclusion, Figure 58 gives the light curves of RR Lyr at different wavelengths (Hardy, 1955).

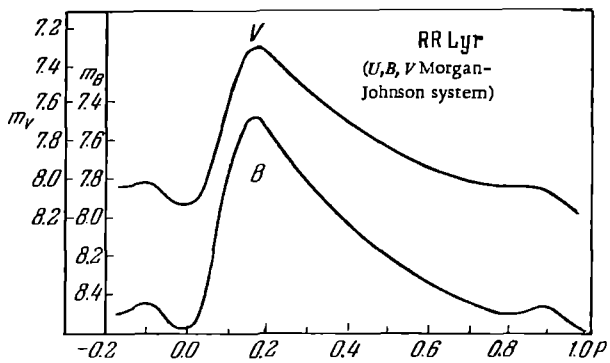


FIGURE 58

2. Spectrum variation curves. The spectra of RR Lyrae stars are of earlier type than those of the long-period cepheids with the shortest periods. They range from late B subtypes to late F subtypes. These spectra are furthermore "anomalous." Contradictory results are obtained from hydrogen lines and metallic lines.

The spectrum variation curve on the whole follows the light variation curve. The earlier spectrum is observed at the maximum, and the later spectrum at the minimum of the light curve. For RR Lyr, for instance, the spectral type varies from A7 to cF4 according to the hydrogen lines (Morgan, 1953) and from A2 to A8 (Munch and Terrazas, 1946), from A2 to A8, F1 (O. Struve, Blaauw, 1948), or from B9.9 to A6.2 (Becker, 1940) according to the metallic lines.

Stebbins's six-color photometry fixes the spectral type of RR Lyr as A7 at the maximum and cF5 at the minimum. The spectrum of RR Lyr also shows bright H and K Ca II lines and hydrogen emission lines (H_α). The emission is observed near the phase + $0^P.037$ on the ascending branch, 22 and 60 min after the median phase. The flare persists for 1.4 hrs and is associated with the secondary period of the star. Moreover, these lines split at the beginning of the rise branch, showing a higher intensity long-wave component, whereas at the end of the rise branch the intensity of the short-wave component is higher.

Comprehensive classification of RR Lyrae stars is a more difficult undertaking than the classification of long-period cepheids. The most detailed classification was carried out by W. Iwanowska at Torun (1952) using the data of Yerkes Observatory (19 stars). She used a two-prism quartz spectrograph with Schmidt camera (170 Å/mm dispersion). The spectrograms show distinct H_γ, H_δ, H_η, K Ca II and other lines. In this study, the period—spectrum relation was established for the first time for RR Lyrae stars. Moreover, this proved to be a two-valued relation, as we see from Figure 59 (the vertical bars indicate the amplitude of spectrum variation).

Most of the points for individual stars fall on the top curve, which corresponds to earlier spectra; six stars, mostly of short periods, fall on the lower curve, which corresponds to later types (BC Eri, UZ Leo, TZ Aur, AR Per, RY Com, BB Pup).

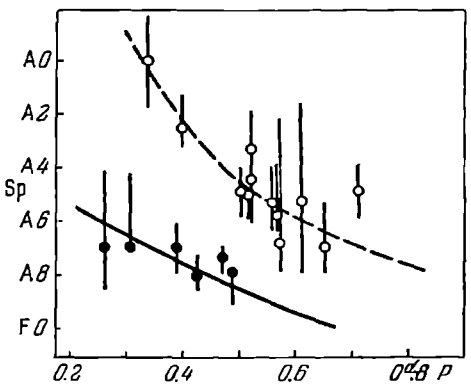


FIGURE 59

The distance between the two curves is 3—4 subtypes. It is remarkable that the stars of the lower sequence, except RY Com, lie closer to the galactic plane; with the exception of BB Pup, they all have lower radial velocities than the stars of the upper sequence. The lower sequence can be tentatively classified as Population I, and the upper sequence as Population II. This classification will be justified in what follows. Stars of sequence I are typically of shorter periods ($0^d.3 - 0^d.4$). It is these stars that, according to Kukarkin, show a more pronounced concentration to the plane of the Galaxy. Population I includes the variable with the shortest known period, SX Phe (HD223061, $P = 0^d.55 = 80$ min) (Eggen, 1952).

According to W. Iwanowska, these two sequences of the spectrum vs. lg P relation for RR Lyrae stars constitute an extension of the corresponding effect for the long-period cepheids. The two populations of RR Lyrae stars (I and II) also show definite concentration toward the galactic plane.

The long-period cepheids of Population II (W Vir stars) are considered in the next section.

Population I and Population II RR Lyrae stars reveal considerable difference in morphological and physical characteristics. Population II stars are poorer in hydrogen and heavy elements. The continuous spectra, the Balmer lines, and the lines of Sc, V, Ti, Eu, Pr, and other elements are weaker in Population I stars. The metallic lines, conversely, are relatively strong. On the whole it seems that the spectra of RR Lyrae stars are similar to the spectra of high-velocity galactic stars.

Figure 60 shows the variation of the spectrum (MKK system) for RR Lyr (O. Struve, Blaauw, 1948) according to two groups of lines: metallic and H, K Ca II lines (cycles 0—13 and 64—74). The earliest spectrum is A2 at $0^P.35$ and the latest is A8, F1 at $0^P.00$. The phases of the minimum and maximum radial velocities are $0^P.40$ and $0^P.00$, respectively.*

* Recently, extensive work on the determination of medium-dispersion spectra (430 Å/mm for H_γ) of some 100 RR Lyrae stars was carried out by G. Preston (1959). RR Lyr stars, like the long-period cepheids, show 1) different spectra in hydrogen lines and metallic lines; 2) marked variation of the spectrum at the minimum (but not at the maximum) with variation in the period. An even more remarkable result is the considerable inhomogeneity of the RR Lyrae stars. A good characteristic is provided by the difference $\Delta S = 10\,[(\text{Sp H}) - (\text{Sp Ca II})]_{\min}$ as a function of Sp Ca II. For types *a* and *b* (Bailey, 1902), $\Delta S = 0-10$ and $(\text{Sp Ca II})_{\min} = \text{F5—A5}$. For type *c* we have 0—10 and F0—A0 and for the group with $P < 0^d.2$ (lower luminosity stars with characteristic galactic rotation, intermediate between the "halo" stars and the "spiral" stars of the Galaxy), $\Delta S = 0-2$ and $(\text{Sp Ca II})_{\min} = \text{F1—A5}$ (but with higher dispersion, as for the preceding group).

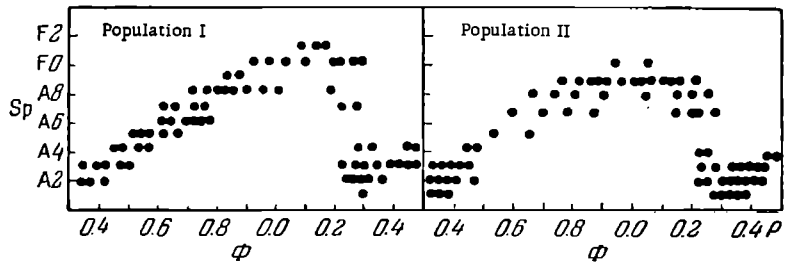


FIGURE 60

3. Radial velocity curves. Complete radial velocity curves covering all the phases are available only for RR Lyr and W CVn. As for the long-period cepheids, these curves are approximately a mirror image of the light curve. According to O. Struve and Blaauw (1948), the velocity curve of RR Lyr, like its light curve, varies with a period of 42^d (75 cycles). The phase shift (retard or advance) of the largest negative radial velocity reaches + $0^p.024$, i.e., it is virtually of the same order of magnitude as the phase of the light maximum. R. Sanford in 1939 studied the radial velocity curve of RR Lyr in great detail using high Coudé dispersion. Back in 1928 he had shown that the amplitude of radial velocities obtained from the H_γ line is greater than the amplitude from the metallic lines. According to the latest studies, the radial velocity amplitude from metallic lines is the smallest, namely 59 km/sec. The hydrogen lines H_8 — H_{12} and H_β, H_γ, H_δ give amplitudes of 71 and 77 km/sec, respectively, the H and K lines of Ca II give 94 km/sec, and H_α line gives the largest velocity amplitude, 112 km/sec.

The hydrogen lines thus give a larger amplitude, which increases for lower quantum numbers.

The phase of the median radial velocity of RR Lyr varies linearly compared to the theoretical value, not unlike the median phase on the ascending branch (Walraven). Figure 61 gives the radial velocity curves according to Sanford. The phases are reckoned from −76 km/sec on the descending branch or, alternatively, from the median visual magnitude on the ascending branch. The dashed curves in Figures 61b, c, d, f are the radial velocity curves from metallic lines. The behavior of the radial velocity curve from the H and K Ca II lines and from the H_α line (solid curves) indicates that the atmospheric waves in RR Lyr suddenly begin moving outward with maximum expansion velocity at the light maximum. After that the layers fall down with maximum compression velocity at the phase of the minimum. The velocity amplitudes from H and K lines are greater than from the metallic lines, at variance with the results of Struve and Blaauw.

Figure 61e gives the velocity variation curve from two series of metallic lines (1 and 2) taken at a large interval of time. The velocities at the minimum are clearly different. The mean shift of the bright H_α line in the spectrum of RR Lyr is −99 km/sec. The maximum velocity difference between the doublets at the time when the two components are resolved is −76 km/sec.

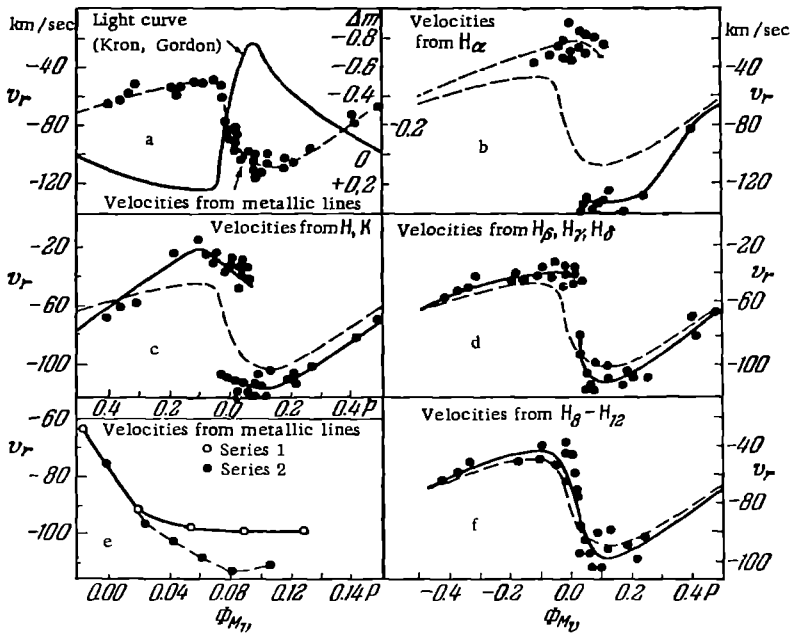


FIGURE 61

The velocity curves of RR Lyr, and apparently also those of other related stars, are thus highly complex. Nevertheless, as for the long-period cepheids, we can calculate the pulsation velocity $\frac{dR}{dt}$ and the change in the star's radius from these curves. The results are shown in Figure 62 (Stebbins, 1953).

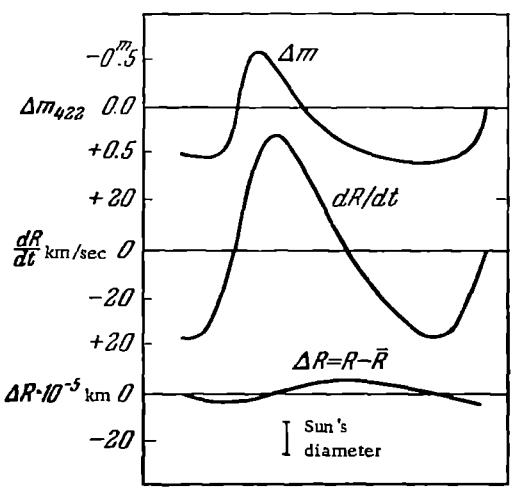


FIGURE 62

4. Surface acceleration curves. Numerical differentiation of the highly complex and variable radial velocity curves of the RR Lyrae stars generally gives uncertain results. The surface accelerations obtained by the method previously outlined for the long-period cepheids are therefore also uncertain. However, the shape of the radial velocity curves suggests that the general trend of surface acceleration should be the same as for the long-period cepheids. The accelerations for RR Lyr range from 40 to 1000 cm/sec^2 (at the maximum).

5. Line intensity curves. In this respect the spectra of RR Lyrae stars have been little studied, except the spectrum of RR Lyr itself, which is given in Figure 63 for wavelengths near the H_α line.

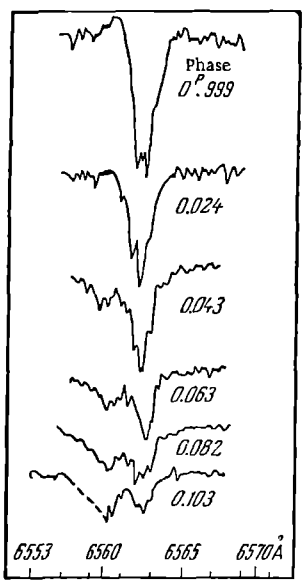


FIGURE 63

At certain phases the lines are indeed split. According to Savedov (1952), the lines are asymmetric, and the skewness is different at the maximum and the minimum. In Savedov's opinion this apparently proves that RR Lyr is a pulsating star.

In the overall spectrum of RR Lyr and apparently other related stars, the changes are not particularly significant. The most prominent changes are observed in the Fe I 4482.2 blend. A violet emission satellite of H_α is observed between $0^P.016$ and $0^P.020$ with average epoch of $-0^P.006$, i.e., very close to the phase of the median magnitude on the rise branch of the light curve. However, the results obtained by different observers in different periods and cycles are somewhat different.

The splitting phase of the H_α absorption line follows immediately the phase of appearance of the emission H_α line. Probably this emission line gradually (in phase) moves to the center of the absorption H_α line and creates an imprssion of a doublet. Besides hydrogen lines, the spectrum of RR Lyr also shows the lines of C I, Fe I, Ti II, V II, Sr II, Ca I, Ca II, Mg II, Si II, and others. In general, the hydrogen lines in the spectrum of RR Lyr are somewhat weak. On the other hand, in stars with $P < 0^d.4$ these lines are apparently of normal intensity at the maximum (CY Aqr, $P = 0^d.06$). At the minimum, they are somewhat weaker than the corresponding Ca II and metallic lines.

The most comprehensive spectrophotometric study of the spectrum of RR Lyr (from the data of O. Struve, dispersion 40 Å/mm) was carried out by Pels-Kluyver in 1954, who used the method briefly described in the previous section in connection with the spectrum of η Aql. The growth curve was plotted for neutral (31 lines) and ionized (31 lines) elements separately.

The growth curve was used to determine the eddy velocity, which was found to be 5.4 km/sec on the average. The variation of excitation and ionization temperature was calculated (from 5250 to 6560°K and from 6960 to 7950°K, respectively); the mean electron pressure, density, surface acceleration and other characteristics were also determined. The variation

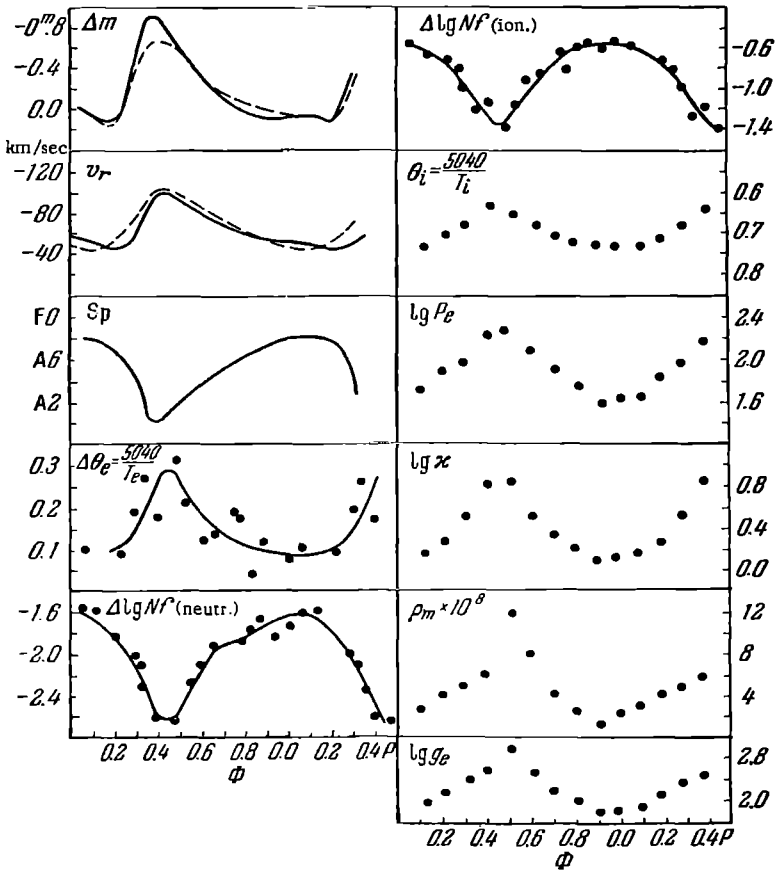


FIGURE 64

TABLE 10

Data \ Star	η Aql	δ Cep	S Sge	T Vul	RR Lyr
P (days)	$7^d.176678$	$5^d.366306$	$8^d.382172$	$4^d.435578$	$0^d.566837$
m_M, m_m	$3.69—4^m.40$	$3.78—4^m.63$	$5\ 8—7^m.0$	$5.87—6^m.75$	$7.1—8^m.0$
ΔV (km/sec)	30	20	18	17	30
Sp (HD)	F6—G4	F5—G2	cF3—cG3	cA9—cG1	A2—F0
$\overline{M}_v$	$-2^m.7$	$-3^m.7$	—	—	$-0^m.4$
$\mathfrak{M}$ (g)	$1.35 \cdot 10^{34}$	$1.78 \cdot 10^{34}$	—	—	$0.646 \cdot 10^{34}$
$\overline{R}$ (cm)	$2.55 \cdot 10^{12}$	$3.82 \cdot 10^{12}$	—	—	$0.457 \cdot 10^{12}$
V_{turb} (km/sec)	4.0	5.0	7.5	6.2	5.4
$\Delta\theta_e = 5040/T_{ex}$	0.21	0.10	0.11	0.06	0.19
$\Delta\theta_i = 5040/T_i$	—	0.17	—	—	0.09
$\Delta \lg P_g$	1.14	1.05	0.4	0.4	0.7
	2.9	10	29	29	40
Dispersion (Å/mm)	20	13 (10)	28	28	67
Number of spectra	M. Schwarzschild et al., ApJ 108, 207, 1948	T. Walraven, Publ. Amst. Inst., 8, 1948	R. Lust-Kulka, Z. f. Ap., 33, 211, 1954	R. Lust-Kulka, Z. f. Ap. 33, 211, 1954	H. A. Pels-Kluyver, BAN, 453, 1954
Author and source					

curves of these parameters, together with light and radial velocity curves (for two phases of the cyclic period P_g) are shown in Figure 64.

Table 10 lists comparative data for a number of long-period cepheids and for RR Lyr. The spectrum of RR Lyr on the whole shows weaker lines at the maximum (A2 spectrum) than at the minimum (A8, F1), which is quite normal. The same tendency is observed for the long-period cepheids.

§ 44. **W Virginis stars.** A detailed study of the long-period cephieds has shown that W Vir stars can be classified as a separate group. Some of their characteristics markedly differ from the characteristics of normal long-period cepheids. The latter are conventionally designated Population I, whereas W Vir stars are designated Population II. W Vir stars have much in common with RR Lyr stars, although their spectra are of markedly later type. Indeed, they show emission lines and doublets, which lead to discontinuous jumps in the radial velocity.

Population I cepheids have periods from 1 to 50 days, a light curve and spectrum which change with P, high concentration toward the galactic plane, small peculiar motion, etc.

Population II cepheids (W Vir and TW Cap stars) constitute a unique group which in fact satisfies only few or perhaps only one of the criteria listed for Population I cepheids. The light curves of Population II cepheids markedly differ from the corresponding average curves of Population I. Population II cepheids occur as an exception in globular clusters. These cepheids lie in higher galactic latitudes, at greater distances from the galactic plane. Their spectra as a rule contain emission lines. Most of them show large peculiar velocities. For example, the velocity of W Vir is virtually equal to the velocities of the fast stars in the Galaxy on the whole. Only the light curves of Population II cepheids have been studied so far.

Among the known cepheids with periods of 1 to 50 days, a few dozen lie at distances of over 500 parsecs from the galactic plane. Their periods are between 2.21 and $28^d.58$. Most of these cepheids, however, have periods between 15 and 20 days. This is reminiscent of the frequency of periods in cluster variables, but definitely not in galactic field cepheids. Most of these stars have variable periods, large peculiar velocities, and unique light curves. Figure 65 shows the photographic light curve of W Vir (S. Gaposchkin), the radial velocity curve (A. Joy), and the spectrum variation curve (C. Payne-Gaposchkin). On the latter circles mark absorption spectra and dots spectra with emission lines. The velocity curve is a smoothed average curve. The true variation of radial velocity, obtained by Sanford (1952) from measurements with high Coudé dispersion, is shown in Figure 66. It clearly reveals the splitting and broadening of lines in the spectrum of W Vir at various phases.

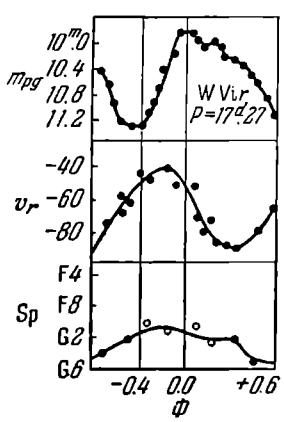


FIGURE 65

For comparison the same figure gives the photoelectric light curve from the observations of Kron and Gordon. The splitting of lines in W Vir is observed near the phase of light maximum

(Sanford, 1952). During this phase, one of the line components still shows a high expansion velocity of the atmosphere (≈ -25 km/sec). After that this component shifts toward the red (to wavelengths corresponding to maximum compression velocities of some +25 km/sec); it does not disappear at the maximum and actually persists for some time after the light maximum. On the other hand, the first component showing the maximum expansion velocity reappears somewhat before the light maximum, etc.

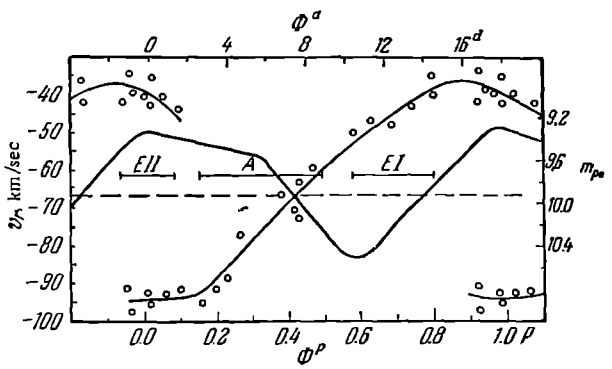


FIGURE 66

Thus, the two velocity sequences and the corresponding line components begin slightly before the light maximum in W Vir. The first red component abruptly vanishes soon after the maximum, whereas the second, violet, component develops continuously until the next maximum, gradually shifting to the position of the red component. It then disappears soon after the maximum. The second component reappears soon before the last maximum in the form of a violet component (negative radial velocity relative to the center of mass of the system, which corresponds to expansion) and continues developing until the next maximum. The figure also marks the horizontal intervals in which a pure singlet absorption spectrum (A), singlet absorption spectrum with bright hydrogen lines (*E* I) and doublet absorption spectrum with bright hydrogen lines (*E* II) are visible. It is significant that the doublets Sr II 4077 and 4215 Å and some high-intensity Ti II lines in the spectrum of W Vir show highly pronounced broadening and shift near the maximum in the spectra of other cepheids as well (Jacobsen, 1949). However, this may be an apparent effect associated with greater blending of lines because of careless application of low-dispersion spectrographs; alternatively the corresponding stars may be too weak. The best objects are the bright cepheids which are accessible to high-dispersion spectrographs. SV Vul ($P = 45^d.21$) apparently gave positive results for line broadening and shift.

Note that in the spectra of some W Vir stars which show no hydrogen emission lines, the absorption lines of hydrogen are exceedingly weak (e.g., V 410 Sgr).

W Vir stars have large radial velocities. Their peculiar velocities are also large. Both factors are generally correlated.

Some of the W Vir stars lie close to the galactic plane, e.g., SZ Mon ($P = 16^d.38$). Its light curve is highly unusual, the hydrogen lines are weak, with possible emission. The same also applies to BH Oph with $P = 11^d.05$. Its spectrum shows weak hydrogen lines (but the light curve is uncertain). Other interesting stars include UZ Sct ($P = 14^d.75$, weak hydrogen lines, normal light curve of small amplitude), AA Ser ($P = 17^d.16$, weak hydrogen lines, possible traces of emission, light curve uncertain).

Population II stars were identified in 1951 by C. Payne-Gaposchkin. These stars are close to the galactic plane. A number of authors are of the opinion that $\varkappa$ Pav ($9^d.1$), Y Oph ($17^d.1$), and AA Ser ($17^d.2$) are intermediate between Populations I and II.

§ 45. **β Canis Majoris stars.** Although not proper cepheids, these stars have some characteristics which relate them to the cepheids. It has been further established that the variations in both types of stars are due to pulsations. β CMa stars are of early B subtypes, although A and even F0 stars are sometimes observed. They show rapid, generally periodic fluctuations of radial velocity with periods from 0.1 to 0.4 days. This effect was discovered by Frost (1902) at the Yerkes Observatory for β Cep ($P = 4^h34^m.2$, $\Delta v_R = 34$ km/sec).

In some cases the radial velocity variation can be represented by a linear relation (e.g., β Cep), but in other cases such representation is possible (12 Lac and others). The radial velocity curve is often variable in shape, amplitude, and phase. Light variation, though insignificant ($0.05 - 0^m.15$), is almost in phase with the radial velocity variation and has the same period. For β Cep these variations were discovered photoelectrically by P. Guthnick in 1913 at the Berlin—Babelsberg Observatory. Prior to this discovery, β CMa stars were regarded as spectroscopic binaries with a small inclination of the orbital plane.

In case of β Cep, the maximum magnitude leads by a mere $0.14 - 0^p.17$ the phase of maximum negative radial velocity. α Lyr is possibly also one of the β CMa stars; in 1930 Belopol'skii discovered radial velocity fluctuations for this star. Later its brightness was also found to vary along a curve which was the "mirror image" of the radial velocity curve.

The absorption line widths and intensities in β Cep stars are slightly variable with phase. The element γ — the velocity of the "center of gravity" of the system — is also variable (β Cep, 12 Lac). Integration of the radial velocity curves gives small variations of the radius — about $5 \cdot 10^9$ cm. For cepheids this variation is generally larger by 2 orders of magnitude.

A study of the differential shifts of lines in the spectrum of β CMa carried out by W. Struve and van Hoof (1953) has shown that the shifts are different for lines with different excitation potentials. A similar effect was observed for 16 Lac.

The velocity curves from hydrogen and other absorption lines (O II, Si II, and others) are similar, though showing a relative displacement (the hydrogen curve leading). A similar effect was found for BW Vul (HD 199140), σ Sco and 12 DD Lac. It was observed, however, only on the descending branch of the velocity curve.

Figure 67 shows the light and radial velocity curves of σ Sco, a β CMa star (Struve, Williams). We see from the figure that these curves are indeed reminiscent of the corresponding curves of the cepheids. However,

the phase relation is slightly different, and actually shows a better fit with the pulsation theory. This, in the author's opinion, is associated with the smaller extent of the atmospheres in β CMa stars compared to the atmospheres of the cepheids. Like cepheids, the β CMa stars satisfy the period—spectrum relation, as was recently established by McNamara (1953). The corresponding data are listed in Table 11 (in the MKK system). For these stars, unlike cepheids, a larger period corresponds to an earlier spectrum.

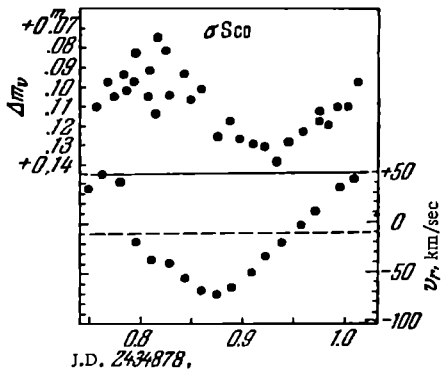


FIGURE 67

The period—spectrum relation for β CMa stars with periods less than 5 hrs is confirmed by the period—color relation discovered for these stars in 1955 by McNamara and Williams. This relation in the $B-V$ Lick system has the form $(B-V)_0 = CI_0 = -0^m.143 - 0.653\,P^d$. For β CMa stars with two fluctuation waves, Table 11 gives only the main component (e.g., β CMa, σ Sco, 12 DD Lac, β Cep, and ν Eri).

TABLE 11

Star	Absolute magnitude	Spectrum (MKK)	Period	
β CMA	$-4^m.7$	B1 II—III	$0^d.25002246$	6^h00^m
σ Sco	—4 .3	B1 III	0 .246835	5 55
ξ^1 CMa	—4 2	B1 III—IV, B2 III	0 .2096	5 02
BW Vul (HD 199140)	—4 .1	B2 III	0 .201028	4 49
12DD Lac	—4 1	B2 III	0 .193089	4 38
β Cep	—4 .1	B2 III	0 .1904844	4 34
ν Eri	—4 .1	B2 III	0 .1735089	4 10
16 Lac	—3 .3	B2 IV	0 .16917	4 04
δ Cet	—3 .3	B2 IV	0 .16122	3 52
γ Peg	—3 .0	B2,5 IV	0 .1517	3 38

The work of O. Struve and his school led to the identification of 10 β CMa stars (they are all listed in Table 11) and helped to sift out all the doubtful cases. Their work has further shown that β CMa stars are in all probability pulsating variables. This, in particular, was confirmed by the color variation of BW Vul (HD 199140) discovered in 1952—1953 and of β Cep. β CMa stars are slightly redder at the minimum and bluer at the maximum. For example, the temperature of 12 Lac at the minimum is some 1000°K lower than at the maximum. Like the cepheids, β CMa stars, e.g., β Cep, show a phase lag of light maxima and minima when passing from the ultraviolet to the infrared spectrum.

In 1918 Anroto noted that the line intensities in the spectrum of β Cep varied with a period of 6^h2^m, which did not fit the period of radial velocity variation.

In the spectrum BW Vul, the H_γ and H_δ lines are enhanced at the minimum, which is a result of the decrease in temperature.

According to McNamara (1953) and also Blaauw and Savedov (1954), the period—luminosity relation is applicable to the β CMa stars, as well as to cepheids. This relation for the ten stars of Table 11 is schematically shown in Figure 68. The absolute magnitudes were derived from the spectral lines and from the motions of a number of stars. Larger periods correspond, as for the cepheids, to higher absolute magnitudes. Petrie in 1954 found $M_v = +0^m.4 - 18^m.1 \lg P$.

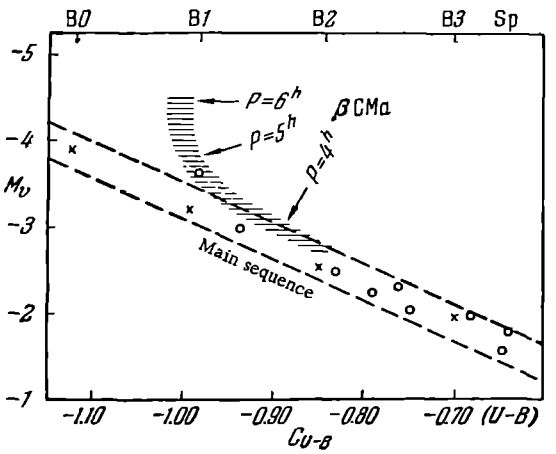


FIGURE 68

A highly important effect observed in some β CMa stars is the superposition of two periods, which leads to cyclic variations (beats). This effect is observed in five out of the ten β CMa stars listed in Table 11. The difference in the periods for β CMa is only two minutes, but this is enough to produce a beating cycle of 49 days. For ν Eri the main period is 4^h10^m with radial velocity varying by 49.0 km/sec. The secondary period is 4^h16^m with radial velocity varying by 22.0 km/sec. The resulting beating period is about 7 days and the radial velocity amplitude is 71.0 km/sec. The light curve gives amplitudes $\Delta m_1 = 0^m.114$, $\Delta m_2 = 0^m.067$, and a total amplitude $\Delta m = 0^m.181$ with minimum value of $0^m.047$. The beating cycles of the light curve are somewhat shifted: the light maximum is observed when the velocity is zero (on the descending branch of each velocity cycle). The velocity amplitude ratio $\frac{(2K_2)}{(2K_1)}$ is equal to the magnitude amplitude ratio $\frac{\Delta m_2}{\Delta m_1}$.

The results and the figures presented in this section have been borrowed from O. Struve's review (1955), where the correlation between $2K$ and Δm is also discussed.

Integration of the radial velocity curves for β CMa stars, as for cepheids, gives the curves of surface variation. Subtracting the corresponding surface component Δm_R from the total light curve, we obtain the amplitude

associated with temperature variations $\Delta m_{T,\lambda}$ (since $\Delta m_\lambda = \Delta m_{T,\lambda} + \Delta m_R$). The corresponding light curves are in phase at all wavelengths. The amplitudes of these curves correspond to an average temperature of 23,000°K. The results give $R - \overline{R} = 5.24 \cdot 10^4$ km $= 0.075\ R_\odot$. The amplitude of the surface light curve, i.e., $\Delta m_R = f(t)$, is taken to be $0^m.036$, so that we should take $\overline{R} = 9.0\ R_\odot$. Then for $T = 23{,}000$°K, $M_v = -4^m.2$ for β Cep. According to Morgan and Keenan (1951), $M_v = -4^m.1$ from the shape of the spectral lines.

In solar units $M_{pv} = M_\odot - 5 \lg \overline{R} + j$, where j is the surface brightness in stellar magnitudes ($M_\odot = +4^m.73$, $T_{e.\odot} = 5700$°K, $T = 23{,}000$°K, B1, $\lambda = 5400$ Å), and thus $\lg j = -4^m.21$. The good fit between the M values calculated from the pulsation hypothesis and those measured directly supports the validity of the theory. This, however, is not always true, especially for stars with two periods. Some of these develop a "hump" on the rise branch of the light curve or the descending branch of the radial velocity curve. Only in low-dispersion measurements this effect is inconspicuous and the lines only appear somewhat more diffuse. High-dispersion observations reveal three line components, although only two are visible at a time.

The splitting times according to Deich are marked in Figure 69 by vertical arrows. We see from the figure that the radial velocity curve of BW Vul has discontinuities. The resolved line components change their intensity very rapidly. The red component disappears at $0^p.697$, and the violet (intermediate) component appears near $0^p.674$, reaching a maximum intensity near $0^p.709$ and vanishing near $0^p.740$. At $0^p.713$ an extreme violet component appears, which progressively increases in intensity and at the same time shifts in the long-wave direction, revealing positive velocity; it thus corresponds to the ascending branch of the radial velocity curve.

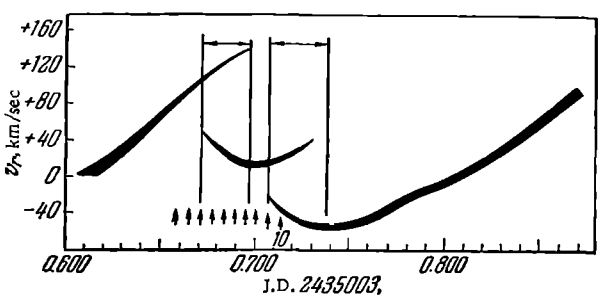


FIGURE 69

The lines in the spectrum of 12 DD Lac are also split (Si III, for example), and the radial velocities change discontinuously. Similar effects are observed for σ Sco. The spectra of these three stars give identical intensity ratio of the H and He lines to the Si III and O II lines, although the former show a phase lag of a few minutes.

We see that the variations in β CMa stars follow a highly complex pattern, and the details cannot be explained within the framework of any of the current theories.

In conclusion note that semiregular RV Tauri and other variables adjoin the sequence of cepheids and β CMa and W Vir stars. These semiregular variables are characterized by light curves with alternating deep and shallow minima with periods from 0.2 to a few hundredths of a day; their spectral types are G, K. These objects as if constitute a link between the periodic variables — cepheids — and the long-period variables.

§ 46. General characteristics of the cepheid population. 1. Distribution of cepheids over the sky. The distribution of long- and short-period cepheids in galactic latitude is highly peculiar. The long-period cepheids show considerable concentration toward the galactic plane, whereas the short-period variables (RR Lyrae stars) are conversely distributed over all galactic latitudes. Various considerations, in particular kinematic arguments, thus led to the division of the entire cepheid population into two groups, as we have mentioned before. Population I cepheids are assigned to Badde's Population I stars, and Population II cepheids to galactic Population II. Kukarkin and Parenago divided the strs into a disk component (roughly corresponding to Population I stars) and a halo component (roughly Population II). Various intermediate systems also exist. Population I includes large quantities of interstellar dust (galactic spiral). Population II (elliptical nebulae, the "halo" of spiral galaxies, globular clusters, etc.), on the other hand, is virtually dust-free. Occasional groups of long-period W Vir stars at high galactic latitudes with mean periods from 1 to 50 days peaked between 15 and 20 days are Population II stars which belong to the halo component of the Galaxy. Some RR Lyrae stars are morphologically Population I stars. Most of the RR Lyrae stars, however, are Population II stars, just like the long-period cepheids are mostly Population I.

W Vir stars are found in globular clusters together with RR Lyrae stars. The brightest of these stars were observed spectroscopically by Joy (M3) in 1949. Their spectra show bright lines (in stars of the same period as W Vir). Similar long-period cepheids, again together with RR Lyrae stars, are observed in the Small Magellanic Cloud, which is mainly composed of Population II stars. On the other hand, the Large Magellanic Cloud, composed mainly of Population I stars, also contains RR Lyrae stars with periods of 0.647, 0.747, and $0^d.913$, whose period—luminosity relation (see below) identifies them as Population I stars (Dartayet and Dessy, 1952). These stars are $1^m.6$ brighter than the RR Lyrae stars of the halo component (the cluster variables) and the stars in the Magellanic Clouds.

Note that RR Lyrae stars, like globular clusters themselves, show a definite concentration in longitude toward the galactic center. This concentration and similarity in their motion led to the assumption that the RR Lyrae stars are in fact escaped cluster variables.

2. The spatial density of cepheids. A first detailed study of the distribution of cepheids in the coordinate z was carried out by Parenago and Grigor'eva in 1937. Later, in 1949, this problem was comprehensively studied by Kukarkin. He determined the log count of cepheids per unit volume, or in other words the logarithmic spatial density lg N.

For RR Lyrae stars he found*

$$\lg N = \lg N_0 + (\mathrm{mod})\, k|\bar{z}| = 7.289 - 0.000219|\bar{z}| \pm 0.125 \pm 0.000016$$

* Unit volume here is the volume of space occupied by RR Lyrae stars of median apparent photographic magnitude 10^m (from 9.5 to $10^m.5$). Unit volume for long-period cepheids is similarly defined.

(here z in parsecs) and for the long-period cepheids

$$\lg N = 1.835 - 0.00991\,|\bar{z}| \pm 0.079 \pm 0.00058.$$

The exponential decrease of density is thus confirmed by observations for all the cepheids.

Note that in statistical studies we should allow for the unequal detection probabilities of cepheids with different amplitudes and apparent magnitudes, cepheids located in dark and bright parts of the Galaxy, etc. In the range of longitudes 235—0—55° (i.e., ± 90° from the center of the Galaxy) there are 290 long-period cepheids, whereas for longitudes 55—235° (± 90° from the anticenter) there are only 173 long-period cepheids. When considering these figures, we must remember that near the anticenter the interstellar medium is much more transparent than in the direction to the galactic center. We thus conclude that the density of long-period cepheids decreases from the galactic center to the anticenter. Other fine features are also observed. For example, van den Bergh found in 1959 that long-period cepheids of short periods (2—3 days) are on the average farther from the galactic nucleus (10—12 kpc) than cepheids of longer periods (7—8 days). Van den Bergh attributed this effect to local differences in the rate of formation of the two cepheid populations. This in its turn may be due to differences in the relative abundance of the heavy elements in the interstellar gas at large distances from the galactic nucleus. This assumption is naturally meaningful only if the cepheids indeed form from the interstellar gas in the Galaxy. This fine feature in the distribution of galactic cepheids is apparently observed in the Magellanic Clouds also.

3. Frequency of periods. The frequency of periods is different for long-period and short-period cepheids. It is furthermore slightly different for free (Galactic) and "bound" (in globular clusters, Magellanic Clouds, etc.) cepheids (Kukarkin, 1949). The short-period cepheids (RR Lyrae stars) are markedly weaker than the long-period cepheids and so far they have not been observed in the spiral arms of the Galaxy. They are known, however, in elliptic galaxies, e.g., in Sculptor and Fornax (W. Baade, E. Hubble, 1939). Baade is of the opinion (1946) that RR Lyrae stars will eventually be discovered in the galactic nucleus and in the spiral arms.

Tables 12 and 13 list the frequency of cepheid periods according to Kukarkin (1949) (nomenclature: N_f "free," solitary cepheids, N_c cluster cepheids, N_{SMC} cepheids in the Small Magellanic Cloud).

The frequencies N_f and N_c in Table 12 (RR Lyrae stars) are essentially different, although the average values of $\overline{\lg P}$ virtually coincide. There is a certain excess of N_f for $\overline{\lg P}$ from −0.3 to −0.4, and an excess of N_c for $\lg P$ from −0.4 to −0.55.

The frequencies N_c also show a secondary maximum (Kholopov, 1944). No cepheids with periods less than $0^d.25$ have been so far discovered in globular clusters. This, however, is partly associated with the observation technique. The above differences are consistent with the fine distinctions in the light curves of free and bound cepheids, despite the overall similarity in crude features.

According to the latest data in Sawyer's catalogue (781 stars), N_c has a minimum near $0^d.45$.

TABLE 12

$\overline{\lg P}$	$N_f=N$	N_c	$\overline{\lg P}$	$N_f=N$	N_c
−1.225	1	0	−0.575	10	6
−1.175	1	1	−0.525	27	23
−1.125	1	0	−0.475	43	32
−1.075	2	0	−0.425	61	59
−1.025	0	0	−0.375	115	24
−0.975	4	0	−0.325	307	669
−0.925	3	0	−0.275	314	156
−0.875	3	0	−0.225	308	145
−0.825	1	0	−0.175	128	73
−0.775	2	0	−0.125	37	21
−0.725	3	0	−0.075	8	8
−0.675	3	0	−0.025	2	2
−0.625	3	9	+0.025	2	0

TABLE 13

$\overline{\lg P}$	$N_f=N$	N_{SMC}	$\overline{\lg P}$	$N_f=N$	N_{SMC}
0.050	6	5	1.150	44	23
0.150	7	37	1.250	31	16
0.250	8	67	1.350	22	7
0.350	13	59	1.450	15	9
0.450	19	90	1.550	11	9
0.550	45	69	1.650	4	2
0.650	57	53	1.750	0	0
0.750	64	34	1.850	1	3
0.850	55	33	1.950	0	1
0.950	26	24	2.000	0	2
1.050	35	21	—		

The frequency curves of Table 13 also differ from one another. The N_f curve has two minima near $\overline{\lg P} = 0.75$ and 1.2. Moreover, $\overline{\lg P}$ for the galactic long-period cepheids is 0.866 and for the cepheids in the Small Magellanic Cloud it is 0.625. This difference cannot be entirely attributed to different visibility conditions (in particular, the detection probability of a long-period cepheid is higher, since by the period—luminosity relation its luminosity is greater; its light variation amplitude and other characteristics are also more prominent). If we allow for this fact, the $\overline{\lg P}$ of free galactic cepheids is found to be 0.606, which is much nearer the corresponding figure for the Small Magellanic Cloud. The light curves of the cepheids in the Galaxy, the Small and Large Magellanic Cloud, and in the spiral nebulae M 31 and M 33 are relatively similar, but the light curves of cluster cepheids are essentially different.

Recently Thackeray and Wesselink (1953) discovered some RR Lyrae stars with mean magnitudes of $18^m.6$, $18^m.7$, and $18^m.7$ in globular clusters NGC 121 (Small Magellanic Cloud) and NGC 1466, 1978 (Large Magellanic Cloud) (3, 21, and 22 stars, respectively).

4. Cepheid motions. The group velocities V_0 of the RR Lyrae stars according to Parenago (1948) are 150± 30 km/sec. The dispersion of their velocities is $\theta = \pm 100$ km/sec. O. Struve reduced in 1950 Joy's values for the radial velocities of RR Lyrae stars and obtained the distribution data listed in Table 14 and 15.

TABLE 14

Range of P values	N	$\overline{P}$	Mean velocity $\overline{v}$ ± mean error, km/sec
$<0^d.20$	7	0.12	44±15.7
$0^d.20$—$0^d.30$	9	0.26	42 10.5
0 .30—0 .40	17	0.34	62 11.5
0 .40—0 .50	35	0.45	78 15.8
0 .50—0 .60	43	0.55	110 15.5
0 .60—0 .70	16	0.64	120 26.9
0.70	6	0.95	80 32.7

TABLE 15

Range of P values	N	$\overline{P}$	Solar motion (velocity) $v_\odot$, km/sec
$<0^d.40$	30	0.27	57
$0^d,40$—$0^d.50$	27	0.45	156
0 ,50—0 .60	34	0.55	200
0 ,60	15	0.68	233

O. Struve has noted that these data bear out Kukarkin's idea that the periodic stars form their own subsystems with characteristic galactic rotation. Pavlovskaya (1953) has been engaged on a detailed study of proper motions of the RR Lyrae stars since 1945. Taking $V_\odot = 134$ km/sec for the velocity of the Sun and $\theta = |V| = 67$ km/sec for the mean "mobility" of stars, she determined the average parallaxes of the RR Lyrae stars:

$$\overline{p_\nu} = 0''.00080 \pm 0''.00008,$$
$$\overline{p_\tau} = 0''.00123 \pm 0''.00016.$$

The motion of long-period cepheids was studied in detail by Mel'nikov (1944) from radial velocity data obtained by Joy.

Table 16 lists the average results. A negative zero point of the period—luminosity relation was used in these calculations ($\Delta M_0 = -0^m.5$).

We see that the "mobility" of the long-period cepheids is virtually independent of their period, which clearly distinguishes them from the RR Lyrae stars. At the same time the mobility of normal stars in the Galaxy increases with decreasing mass and vice versa.

In his study of the long-period cepheids, Mel'nikov established the galactic rotation constant $A = 17.2$ km/sec/kpc, the dependence of the (negative) K term on the period $K = +1.8 - 9.1 \overline{\lg P}$, and the velocity of the Sun. These data are listed in Table 17. On the average, we have from this table $V_\odot = 18.4 \pm 1.4$ km/sec. Four years later, Parenago (1948) found $V_\odot = 18.6$ km/sec. The velocity of the Sun is a function of the period, as we see from Table 17.

TABLE 16

$\overline{\lg P}$	Average velocity $\lvert v \rvert \pm$ mean error, km/sec	N
0.68	11 ± 1	39
0.92	12 ± 1	34
1.28	13 ± 1	22

TABLE 17

$\overline{\lg r^*}$	0.45	1.08	1.68	2.79
$V_\odot$, km/sec	12.2	20.8	14.6	27.6
K, km/sec	— 2.9	— 1.3	— 9.2	—16.1

* Here $\lg r(\text{pc}) = f(\lg P)$.

The analysis of proper motions carried out by Blaauw and Morgan (1954) gave highly accurate data on the motions of 18 long-period cepheids. The average secular parallax was found to be $0''.0120 \pm 0''.002$, which taking for the Sun's velocity 18.4 km/sec (Mel'nikov, 1944) or 18.6 km/sec (Parenago, 1944) gives the normal parallax of $0''.00310$ or $0''.00305 \pm 0''.00058$.

Note that proper motions, especially those of long-period cepheids, largely reflect the biased errors of the particular fundamental system used.

W Vir variables, like most RR Lyrae stars, are Population II stars. They move faster, and their proper motions are therefore more significant. Their mean residual velocities are 40 km/sec (Kukarkin), i.e., a factor of 4 greater than for Population I long-period cepheids.

β CMa stars have been insufficiently studied for a statistical analysis.

5. Period — spectrum relation. This problem was discussed in the preceding for individual cepheid groups. We will therefore only give Parenago's summary curve (Figure 70).

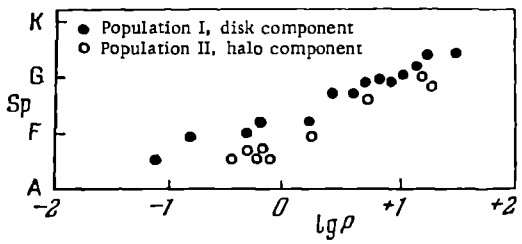


FIGURE 70

We see from the figure that Population II cepheids entering the halo component of the galaxy (marked by light circles) have earlier spectra (for the same period) than the Population I cepheids of the disk component. Figure 71 gives the period—spectrum relation for the entire so-called "large sequence" of variables (Kukarkin, 1948), with some results for β CMa added (McNamara, 1953) (nomenclature: 1 β CMa, 2 RR Lyrae stars, 3 long-period cepheids, 4 RV Tau, 5 semiregular variables, 6 O Cet).

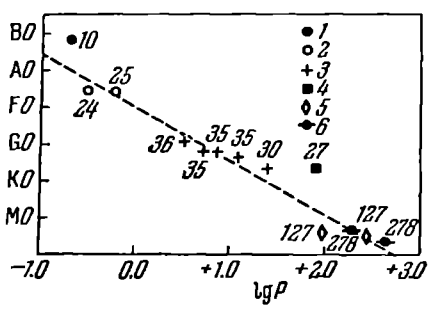


FIGURE 71

Note that Population II cepheids (W Vir stars) have the same characteristics as the cluster variables of comparable periods (e.g., the stars in cluster No. 154, M 3), although their characteristics are entirely different from the characteristics of the classical long-period cepheids of the same period which are assigned to Population I.

6. Period — density relation. This is a highly significant relation, in so far as it enables us to compare the results of observations with the theory and to evaluate the fit of calculated and observed periods. The mass—density relation is satisfied by cepheids of all populations. This relation follows from the theory and from quite general considerations (see below). It has the form

$$P\sqrt{\bar{\varrho}} = \text{const}.$$

Since

$$\varrho \sim \frac{\mathfrak{M}}{R^3},$$

we have

$$\frac{\bar{\varrho}}{\bar{\varrho}_\odot} = \frac{\mathfrak{M}/\mathfrak{M}_\odot}{(R/R_\odot)^3}$$

and

$$\lg \frac{\bar{\varrho}}{\bar{\varrho}_\odot} = \lg \frac{\mathfrak{M}}{\mathfrak{M}_\odot} - 3 \lg \frac{R}{R_\odot}.$$

But

$$\frac{L}{L_\odot} = \left(\frac{R}{R_\odot}\right)^2 \left(\frac{T}{T_\odot}\right)^4$$

and

$$M - M_\odot = -5 \lg \frac{R}{R_\odot} - 10 \lg \frac{T}{T_\odot}.$$

Therefore

$$\lg \frac{\bar{\varrho}}{\bar{\varrho}_\odot} = \lg \frac{\mathfrak{M}}{\mathfrak{M}_\odot} + 0.6 (M - M_\odot) + 6 \lg \frac{T}{T_\odot}.$$

Hence, finally, we get

$$\lg \sqrt{\frac{\bar{\varrho}}{\bar{\varrho}_\odot}} = \frac{1}{2} \lg \frac{\mathfrak{M}}{\mathfrak{M}_\odot} + 0.3 (M - M_\odot).$$

This quantity can be calculated for various periods under the following two assumptions:

(a) $\frac{\mathfrak{M}}{\mathfrak{M}_\odot} = \left(\frac{L}{L_\odot}\right)^{\frac{1}{n}} = f(P)$, i.e., assuming that Eddington's mass—luminosity relation is applicable (this relation was established for normal stars (1926), but never proved for cepheids), and

(b) $\frac{\mathfrak{M}}{\mathfrak{M}_\odot}$ = const, i.e., assuming independence of mass and period (Mel'nikov, 1948).

The computational difficulties stem from the fact that we require the integrated, bolometric luminosities L_b or M_b. They are generally derived from the visual magnitudes M_v using the bolometric correction $\Delta M_b = -\text{B. C.} = M_v - M_b$, which in its turn is computed from Planck's equation or is borrowed from the theory of continuous spectra (neither of these possibilities is particularly reliable). A simplifying factor is that for spectral types near G the bolometric correction is at its minimum and the visual magnitude can be substituted for the bolometric magnitude; this, of course, does not hold for the extreme spectral types of the cepheids (A or K).

Assumption (a) was used by M. Gussov (1927) and C. Payne—Gaposchkin (1930). Payne-Gaposchkin obtained $P \sqrt{\frac{\bar{\varrho}}{\bar{\varrho}_\odot}} = c = 0.12$ (P in days). This

relation was also investigated by Z. Kopal (1939), W. Becker (1940), N. A. Kozyrev (1951), R. Kraft (1953), and others. According to Kozyrev and Kraft, $P\sqrt{\frac{\bar{\varrho}}{\varrho_\odot}} = 0.075$ and 0.05 at the respective zero points of the period—luminosity curve $\Delta M_0 = 0^m.0$ and $-1^m.5$.

Using assumption (b), Mel'nikov (1948) obtained the following inequality (with zero point correction $\Delta M_0 = -0^m.5$)

$$0.049 < P\sqrt{\frac{\bar{\varrho}}{\varrho_\odot}} < 0.093.$$

The dependence of P on $\sqrt{\frac{\bar{\varrho}}{\varrho_\odot}}$ according to the results of various authors is plotted in Figure 72. The data include those of Payne—Gaposchkin and Gaposchkin (1943) for $\Delta M_0 = 0^m.0$, Kraft for $\Delta M_0 = -1^m.5$, and Mel'nikov for $\Delta M_0 = -0^m.5$. Assuming the constant mass hypothesis, Meln'ikov obtained

$$\lg \bar{\varrho} = \text{const} - 2.02 \lg P.$$

This hypothesis was supported by Parenago (1955), and Thiessen in 1956 actually proved the inapplicability of the ordinary mass—luminosity relation to cepheids (see above). Figure 72 shows that both assumptions alternatively satisfy the relation $P\sqrt{\bar{\varrho}} = \text{const}$, although with a different constant. The theoretical calculations of Severnyi (1948) gave the following limits for $P\sqrt{\frac{\bar{\varrho}}{\varrho_\odot}} = c$:

$$0.050 < P\sqrt{\frac{\bar{\varrho}}{\varrho_\odot}} < 0.088.$$

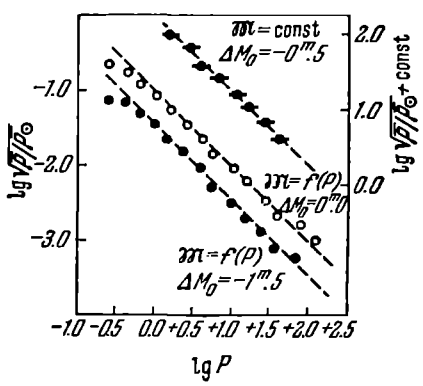


FIGURE 72

For the standard model with specific heat ratio $\gamma = 5/3$, the constant c is equal to 0.0425. The limiting values derived by Severnyi correspond, on the one hand, to a star with a vanishingly small admixture of the heavy

elements and, on the other hand, to a star with a large content of these elements. We see that the constants obtained by Mel'nikov, Kraft, and others comfortably fall within these limits.

Under assumption (a), the masses are generally borrowed from the mass—luminosity relation in Kuiper's form (1939) or, still better, in the form of Parenago and Masevich (1951) for high-luminosity stars (from -1^m to -7^m). Note, however, that for such high-luminosity stars this relation was only verified using few spectroscopic binaries. It is therefore not completely certain and its application to cepheids is highly questionable, especially if we recall Thiessen's result for BM Cas (1956). In this system the mass—luminosity relation apparently breaks down.

In 1939 Z. Kopal carried out a new determination of the constant $P\sqrt{\bar{\varrho}}$ for 88 long-period and 30 short-period cepheids. He found that observational data give a large spread of values, which was attributed to the correlation with the spectral type that he had discovered. Figure 73 plots Kopal's results.

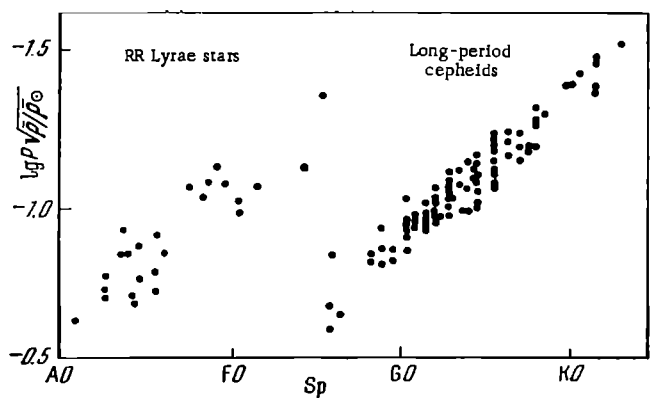


FIGURE 73

We see that there is a distinct gap between long-period and short-period cepheids, which is associated with the discontinuity in the period—spectrum relation.

The constant $P\sqrt{\bar{\varrho}}$ was also revised for the short-period cepheids in the cluster M 3 (Roberts and Sandage, 1954).

The relation between theoretical and observed cepheid pulsation constants $P\sqrt{\bar{\rho}}=c$ for a number of models still fails to give unambiguous results. Under certain limiting conditions for various models the theory gives $0.028 < c < 0.070$.

7. Period — luminosity relation. The period—luminosity relation of some physical variables in general and cepheids in particular is of considerable interest for various purposes. It is particularly important to know the exact cepheid luminosities for comparison of observational data with the results of various theories of variability of cepheids of a given

period. Even greater significance is attached to the mass—luminosity relation as a yardstick for the observed part of the Universe, to be used in measuring the distances of the far galaxies. Indeed, the determination of luminosity, and hence the distance, from the easily measurable period is a most attractive proposition. Alternative methods of distance determination of the distant stars are less reliable.

Unfortunately, it has been recently established that various characteristics of cepheids differ in stellar systems of different ages. The above method of distance determination is therefore now applied only to the Local Group, and even here with some reservations. It is hoped, however, that the cepheid method will be used on a much larger scale as soon as the above-mentioned differences in characteristics have been studied. Detailed analysis reveals a highly promising future. Our interest in the mass—luminosity relation should thus steadily increase.

This most important relation was discovered by H. Leavitt at Harvard in 1908 and finally confirmed in 1912 using Southern Hemisphere observations of the Small (and later the Large) Magellanic Cloud, where the minimum and the maximum stellar magnitudes of cepheids with periods from 1.6 to 66 days were determined (the short-period cepheids — RR Lyrae stars — were discovered much later, and only in small numbers). The apparent magnitudes of the cepheids in the Small Magellanic Cloud differ from the absolute magnitudes by a constant amount (the distance modulus), which is equal to $M-m = 5-5 \lg r = 17^{m}.25$. The apparent, and hence the absolute, magnitudes of the long-period cepheids were found to show an almost linear correlation with the logarithm of the period. This relation between the period and the mean, or median, absolute magnitudes since then became known as the period—luminosity relation.

Approximately at that time, S. Bailey was engaged at the Harvard Observatory on a study of globular clusters, where he discovered mainly short-period cepheids (for which he suggested the name cluster variables). He established that cluster variables with periods from 0.2 to 1.0 days did not follow any period—luminosity relation. The dispersion of their magnitudes about the mean was a mere 0.1 stellar magnitude.

The work of H. Shapley demonstrated the extraordinary significance of the period—luminosity relation for determining the distances of globular clusters and extragalactic systems (primarily M 31 in Andromeda and M 33 in Triangulum). First, however, the zero point of the period—luminosity relation had to be determined, so as to permit converting the apparent magnitudes to absolute magnitudes. This problem was first investigated by E. Hertzsprung in 1913 for 13 galactic cepheids with known proper motions from Boss's catalogue. The secular parallaxes of the group, and hence the distance modulus, could thus be determined. In 1918 a similar analysis was carried out by Shapley using 11 stars with proper motions listed in Boss's catalogue and with known radial velocities. Statistical group parallaxes were used, since the individual distances of galactic cepheids are seldom known: the cepheids as a rule do not appear as components in multiple stars and visual binaries with known parallaxes, nor are they found in Hyades-type moving clusters, open clusters, or other formations whose distances are known from the Hertzsprung—Russell diagram or from other data.

Bailey's observations detected occasional long-period cepheids in some of the globular clusters in the Galaxy. For example, ω Cen was found to contain 5 long-period cepheids, against more than 100 cluster variables. Comparison of the apparent magnitudes of long-period and short-period cepheids in globular clusters, using the period — luminosity relation (from Leavitt's data for the long-period cepheids in the Small Magellanic Cloud; Bailey's data for globular clusters established lack of any correlation for the short-period variables), enabled Shapley to plot the entire period—luminosity curve. In this way the "shape" of the period—luminosity function was fixed for all the cepheids with periods ranging from very short to very long. The visual magnitudes were used. The zero point was fixed from the motions of 13—11 galactic cepheids, i.e., from average parallaxes. Shapley confirmed Hertzsprung's result: he obtained for 11 cepheids $M_v = 2^m.35 \pm 0^m.19$ for $P = 5.96$ days. Once both the shape and the zero point of the period—luminosity relation had been fixed, the short-period cepheids were found to have $M_v = -0^m.3$. The curve enabled the astronomers to determine the distances to objects containing cepheids of any period. Particularly suitable for this purpose are the short-period cepheids, whose luminosity is independent of the period. In these cases we do not have to determine the period and only the median, or the mean, stellar magnitude is required. This method, however, is no longer used, since we have seen that the long-period cepheids in globular clusters are Population II stars, or, more precisely, stars of the halo component of the Galaxy, and are thus physically different from the Population I long-period cepheids of the disk component. There is thus no justification whatsoever to combine these stars into a single curve.

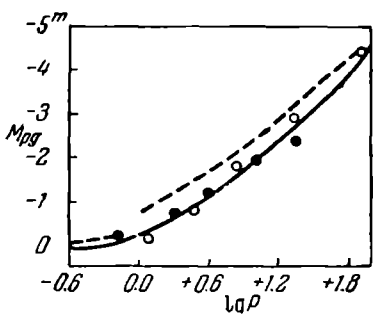


FIGURE 74

In 1930, in his book "Star Clusters" Shapley has finally adopted after a lengthy discussion a definitive zero point of the photographic period—luminosity curve for cepheids. However, he introduced an unjustified correction of $+0^m.25$, which as it turned out had a detrimental effect on the originally excellent result obtained as early as 1918 using a small 11 star sample. Shapley's curve is given in Figure 74 (solid curve). The zero point of this curve according to Shapley is $\Delta M_{s,pg} = 0^m.0$. The luminosity of the short-period cepheids is taken equal to $M_{pg} = 0^m.0$. Dark circles in the figure correspond to average log periods $\overline{\lg P}$, the light circles mark the average M.

As proper motions and radial velocities of more stars became known, various authors tried to revise the shape of the period—luminosity relation and especially its zero point (1923—1939).

All the authors used the proper motions, radial velocities, and trigonometric parallaxes of the cepheids. As a rule the absolute magnitudes $M_{s,pg}$ from Shapley's relation were found to be too low. Later this was linked up with corrections for interstellar absorption, which had been ignored by most of the authors. The significance of these absorption corrections was finally recognized by astronomers only in the 1930's.

In 1944 Mel'nikov revised the question of the zero point of long-period cepheids using radial velocities, proper motions, trigonometric and spectroscopic parallaxes. At the same time he considered the problem of interstellar absorption. After 1930 the zero point had been shifted to $\Delta M_{pg} = -0^m.5$. The differences between the averages $\overline{\lg r}$ and $\lg \bar{r}$, $\lg \bar{p}$ and $\overline{\lg p}$, and also $c=\bar{r}\cdot\bar{p}$ were taken into consideration, a point which proved of considerable significance. Suppose that we are seeking to determine ΔM from the average distance $\bar{r}$ calculated from the equation of differential galactic rotation. Let M_s be the absolute magntide from Shapley's period — luminosity relation, M the true absolute magnitude, ΔM the zero point correction, and a_{pg} a photographic absorption coefficient. Let further ρ be the "photometric" distance (not corrected for absorption) and r the true distance. Then

$$M_s = m + 5 - 5 \lg \varrho,$$

$$M_r = M_s + \Delta M_r = m + 5 - 5 \lg r - \frac{a_{pg} \cdot r}{1000}.$$

Averaging and subtracting we get

$$\Delta M_r = 5\,\overline{\lg \varrho} - 5\,\overline{\lg r} - \frac{a_{pg}\bar{r}}{1000} = 5\,(\lg \bar{\varrho} - \lg \bar{r}) + C_r - \frac{a_{pg}\cdot \bar{r}}{1000},$$

where

$$C_r = 5\,[(\lg \bar{r} - \overline{\lg r}) - (\lg \bar{\varrho} - \overline{\lg \varrho})].$$

Similarly, if the average true parallax $\bar{p}$ is known from proper motions, we have

$$\Delta M_p = 5\left(\lg \bar{\varrho} - \lg \frac{c}{\bar{p}}\right) + C_p - \frac{a_{pg}\cdot c}{1000\bar{p}},$$

where

$$c = \bar{r}\cdot\bar{p}.$$

C_r, C_p, and c can be tabulated for various values of ΔM and a_{pg}. The intersection point of the curves $\Delta M_r = f_1(a_{pg})$, $\Delta M_p = f_2(a_{pg})$ for two groups of stars fixes the zero point correction ΔM and at the same time a_{pg} (since in practice groups of stars with proper motions and radial velocities have entirely different composition). Thus, Mel'nikov obtained $\Delta M = -0^m.5 \pm \pm 0^m.2$ and $a_{pg} = 0^m.91\,\mathrm{kpc}^{-1}$. These results were confirmed by an analysis of the z coordinate of cepheids ($\Delta M = -0^m.5$ for $a_{pg} = 1^m.1\,\mathrm{kpc}^{-1}$) and by the trigonometric parallaxes ($\Delta M = -0^m.4$ for $a_{pg} = 1^m\,\mathrm{kpc}^{-1}$).

Mel'nikov's analysis of the known 156 radial velocities and 86 proper motions of long-period cepheids gave $\Delta M = -0^m.5$ and $a_{pg} = 1^m\,\mathrm{kpc}^{-1}$. This means that the distance of the extragalactic nebulae in Andromeda, Triangulum, etc., should be multiplied by a factor 1.26. This question was also studied in 1944 by Camm and Mineur. Mineur returned to this problem in 1952. Camm obtained $\Delta M = -0^m.11 \pm 0^m.19$, i.e., virtually zero, whereas Mineur found $\Delta M = -0^m.73$ and later $-1^m.3$, which is consistent with Mel'nikov's result as far as the sign is concerned. Kukarkin in 1949 revised the question of the zero-point correction and the shape of the

period—median absolute magnitude relation. After a critical analysis he adopted Mel'nikov's zero point, i.e., $-0^m.5$ (long-period cepheids) and obtained the following relations:

1) for long-period cepheids

$$M_{pg} = -0^m.74 - 1^m.67 \lg P \; (0.00 < \lg P < 0.95),$$
$$M_{pg} = -0^m.35 - 2^m.08 \lg P \; (0.95 < \lg P < 2.00);$$

2) for short-period cepheids

$$M_{pg} = -0^m.17 - 0^m.20 \lg P \; (\lg P < 0).$$

These relations are plotted by dashed curves in Figure 74. They have obvious discontinuities for $\lg P = 0.0$ and $\lg P \approx 1.0$. The above results would appear to confirm the negative zero point, but most astronomers did not realize the significance of this result and up to 1952 continued using Shapley's curve with the zero point $\Delta M_{pg} = 0^m.0$.

In 1952 Baade turned his attention to this question and in 1956 he published a comprehensive review of the entire zero-point dispute.

Having photographed the Andromeda Nebula M 31 with the 200-in. telescope with a correction lens ($F/3.7$), contrary to all expectations he failed to discover any short-period cepheids, even at the maximum. He thus concluded that Shapley's zero point $\Delta M_s = 0^m.0$ of the period—luminosity curve should be moved to $\Delta M = -1^m.5$ for the long-period cepheids, whereas for the short-period cepheids the previous value can be retained.

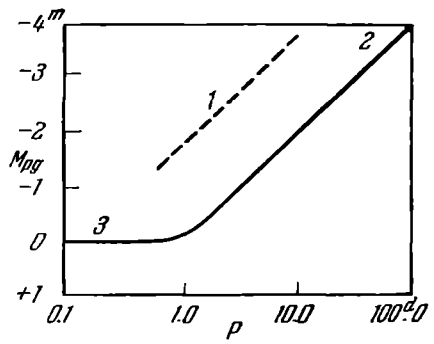


FIGURE 75

This was the second adjustment of the zero point of Shapley's period — luminosity relation after Mel'nikov's study (1944), who obtained $-0^m.5$.

Note that W Vir stars, i.e., Population II long-period cepheids (stars of the halo component), retain their original luminosities according to Shapley's period—luminosity curve with $\Delta M_s = 0^m.0$ even if the zero-point corrections $-0^m.5$ or $-1^m.5$ are adopted.

The entire period—luminosity curve according to Baade thus falls in the first approximation into two separate parts: a branch of Population I cepheids and a branch of Population II cepheids, i.e., cepheids of the disk and the halo subsystem. This is schematically shown in Figure 75 (1 corresponds to Population I long-period cepheids, 2 to Population II long-period cepheids, 3 to Population II short-period cepheids, RR Lyrae stars). If Baade's result receives further substantiation, it will not affect the size of the Galaxy, the distances of the globular clusters, etc., since they are mainly based on the luminosities of the short-period cepheids, which have not changed either. The dimensions of other galaxies and their distances,

on the other hand, should be doubled, as they are based on the luminosities of Population I long-period cepheids with $\Delta M = -1^m.5 = -5 \lg \frac{r(\Delta M = -1^m.5)}{r(\Delta M = 0^m.0)}$.

Baade's results created a forceful impression and all the subsequent authors took it upon themselves to prove (or disprove) the validity of the new zero point $\Delta M = -1^m.5$. A similar tendency became apparent soon after the publication of Shapley's work in 1930, when the zero point $\Delta M_s = 0^m.0$ was first introduced. The need in a positive correction of about $+1^m$ was felt only for the weak cepheids, but it was accounted for by the effect of interstellar absorption. The discussion around the zero-point correction between the years 1930 and 1944 was not particularly effective, and Shapley's zero point $\Delta M_s = 0^m.0$ remained in wide use until 1944.

In 1944 Mel'nikov first established the negative zero-point correction $\Delta M = -0^m.5$. It was not accepted, although subsequent work proved it to be substantially correct (Minuer, Camm, 1944; Kukarkin, 1949, and others). Finally, in 1952 an even larger negative zero point ($\Delta M = -1^m.5$) was advanced and accepted. The preceding discussion refers only to the long-period cepheids of the disk component. A long chain of studies followed intended to verify this zero point. Often they ignored or underestimated various relevant corrections in the calculation of averages, exaggerated the effect of interstellar absorption, etc. A critical analysis of Baade's work led Kholopov (1954) to the conclusion that Shapley's zero point was inapplicable to the short-period cepheids ($\Delta M = +2^m.0$). It was, however, valid for the long-period cepheids of the disk component ($\Delta M = 0^m.0$). In this case the entire situation is naturally reversed: the distances and the dimensions of the galaxies do not change, as they are based on the long-period cepheids, whereas the system of globular clusters is "contracted" by a factor of 2.5 (the corresponding distances are reduced by this factor). In theoretical calculations the mean density of matter in the long-period cepheids should be left as before, and that in the short-period cepheids should be increased by a factor of 15, and so on. Arp (1955) tried to prove that the Population II long-period cepheids in the nearest globular clusters (M2, M5, M10, M13, M15, and ω Cen) with periods of from 1.5 to 67 days and longer fall in their turn into three distinct period—luminosity sequences with different zero points. These sequences are displaced along the horizontal axis by twice the period value, so that only the lowest sequence is a direct continuation of the corresponding curve for RR Lyrae stars. Note that the average luminosities were determined by a somewhat unusual technique: the area of the light curve in intensity units was measured planimetrically to determine the mean apparent luminosity l. This in its turn was used to calculate the mean absolute magnitude M, which was found to be $0^m.2$ greater than the ordinary median magnitude.

Arp is of the opinion that the division of Population II cepheids into three sequences with doubled periods is associated with different pulsation rates of these stars. However, Reddish (1955) rejected this conclusion, based on a small sample of only 19 stars. He showed that 12 out of the 19 stars (although their spectra are known, a prerequisite for the determination of bolometric luminosities) virtually gave a single period—luminosity relation, with marked deviations for two stars only. This was due not to differences in pulsation but to differences in chemical composition. This

point is of considerable significance, since the theory of the internal constitution of stars uses data on mass, luminosity, radius, and chemical composition.

A correct answer (at least as far as the methodology is concerned) to the problem of the zero-point correction of the period—luminosity curve was supplied by Kukarkin in 1954. He observed that the zero point is fundamentally not a single-valued quantity. He suggested that there was a multiplicity of zero points in various stellar systems, which are often populated by stars of different age characteristics.

Baade's result thus only indicates that the zero points of the period—luminosity curves of RR Lyrae stars and long-period cepheids are different, the difference being about $1^m.5$.

Kukarkin showed that a whole range of other morphological characteristics of cepheids are different in different star systems. These include 1) the dependence of the light curve skewness on the period, 2) the dispersion of the absolute magnitudes of RR Lyrae stars ($0^m.15$ in globular clusters and $0^m.6$ in the galactic field, i.e., in bound and free state, respectively), and so on.

In view of the uncertainties in the photometric systems, the cepheid luminosities are best characterized by the integrated or "bolometric" absolute magnitudes M_b. Parenago (1955) obtained the following relations (with some uncertainty in ΔM_b):

1) $M_b = -0^m.49 - 0^m.20 \lg P$ for RR Lyrae stars,

2) $M_b = -0^m.92 - 0^m.51 \lg P$ for Population I long-period cepheids with $P < 9^d$,

3) $M_b = -0^m.14 - 3^m.28 \lg P$ for Population I long-period cepheids with $P > 9^d$,

4) $M_b = -0^m.81 - 2^m.72 \lg P$ average for all periods.

For Population II long-period cepheids Reddish (1955) obtained

5) $M_b = -0^m.07 - 2^m.67 \lg P$.

Note that the period—luminosity relation was also established for RR Lyrae stars in globular clusters, M3 in particular. Attempts are currently being made to calculate the absolute magnitudes of cepheids from the average radii obtained by comparing equal-color points on the radial velocity curves. This was the approach used by A. Rogers (1957). He found $M_{pg} = 0^m.0 - 1^m.74 \lg P$ for $\varkappa$ Pav, γ Oph, W Vir and $M_{pg} = -1^m.6 - 1^m.7 \lg P$ for δ Cep, η Aql, β Dor and l Car. For β Dor, M_{pg} (and therefore R) is relatively high compared to its period. These are the so-called pulsation parallaxes, previously calculated by Kipper. Their applicability has been discussed by various authors.

In conclusion of this subsection let us consider a direct method for assessing the applicability of the zero point. Eggen (1951) used this method for δ Cep A and its companion δ Cep B. The companion, as we have noted before, is separated 40" from the primary and has $m_{pg} = 6^m.01$, $CI = C_p = -0^m.12$, Sp = B8—B9. According to Herbig, this companion is reminiscent of the blue dwarfs in Pleiades. For these stars, for $C_p = -0^m.12$ the absolute magnitude is $M = +0^m.55$ (Eggen, 1950). The distance modulus for δ Cep A inferred from its satellite is therefore $m - M = 6^m.01 - 0^m.55 = +5^m.46$. Hence at the median point of the light curve ($m_{med} = \frac{m_M + m_m}{2} = \frac{3^m.90 + 5^m.00}{2} = 4^m.45$) we have for δ Cep A $M = -1^m.01$. Shapley's period—luminosity curve, on the other hand, gives for δ Cep A $M = -1^m.42$.

It thus follows from these data that $\Delta M = -1^m.01 + 1^m.42 = +0^m.41$, i.e., a positive correction is obtained for this long-period cepheid. Remember, however, that the method is inaccurate, since the position of δ Cep B in the Hertzsprung—Russell diagram is such that a slight change in C_p leads to a large variation in luminosity.

We see that these methods for the determination of ΔM lead to highly contradictory results.* Yet the problem of the exact zero-point correction is of considerable significance, since it is related to the determination of interstellar and intergalactic distances. Recently some authors tried to link this problem with the time scale and the doubling of intergalactic distances, which led them to associate the new zero point $\Delta M = -1^m.5$ with unfounded arguments purporting to explain the evolution of the entire Universe from a single point.

The multiplicity of zero points of the period—luminosity relation even for the galactic long-period cepheids causes considerable confusion. Often it is associated with inaccurate introduction of interstellar absorption.

In view of the considerable uncertainty, it seems that we should look only for corrections to the most reliable zero points, and this only after more complete information becomes available on interstellar absorption, normal colors, true radii of cepheids, etc.

For Shapley's photographic curve (1930) the most reliable zero points are the following

(a) the zero point of the long-period cepheids of the disk component, $\Delta M = -0^m.5$ (Mel'nikov, 1944), $\Delta M = -1^m.0$ (Parenago, 1955), and $\Delta M = -1^m.5$ (Baade, 1952);

(b) the zero point of the short-period (and partly long-period) cepheids of the halo component, $\Delta M = 0^m.0$ (Baade, 1952), $\Delta M = +0^m.5$ (Pavlovskaya, 1953), and $\Delta M = +2^m.0$ (Kholopov, 1954).

All the other zero points, taking account of the actual and inherent errors (and not only the errors described by the respective authors), should be close to the above values.

8. Some general regularities for cepheids. A basic characteristic of all the cepheids is the Hertzsprung—Russell (color-luminosity or spectrum—luminosity) diagram. The complete diagram is shown in Figure 76 from Eggen's photoelectric observations (1951). However, this is only a qualitative diagram, since the normal colors of the cepheids in this system differ from those obtained by a number of other authors (as communicated by G. S. Badalyan). In this diagram 1) $M = +0^m.31 - 2^m.52 \lg P$, i.e., a period—luminosity relation with a zero point slightly different from Shapley's zero point, 2) $C_{p,M} = -0^m.036 + 0^m.613 \lg P$, 3) $C_{p,m} = -0^m.12 + 1^m.194 \lg P$.

For normal stars near the Sun, Badalyan's data were used (1950), and for stars in the globular cluster M92 Hoenberg's data (1939) with distance modulus $m_{pg} - M_{pg} = 15^m.0$. The difference in the amplitudes of cepheids of groups A and B is relatively small, and they are combined into one sequence. Cepheids of group C at the minimum are not shown since 1) it has not been proved that they follow the same period—luminosity relation as cepheids of groups A and B and 2) the period—amplitude relation for this group is inconclusive.

* Recently Yu. P. Pskovskii analyzed the radial velocity data and obtained a zero-point correction $\Delta M = -1^m.3 \pm 0^m.3$ to Shapley's curve (1957, 1959), and Mel'nikov showed that the correction $-\Delta M > 0^m.5$ is apparently inconsistent with the value of the pulsation constant e (IAU Congr., Moscow, 1958; Trans IAU, Vol. X, p. 701, 1958, Cambr. Univ. Press, 1960).

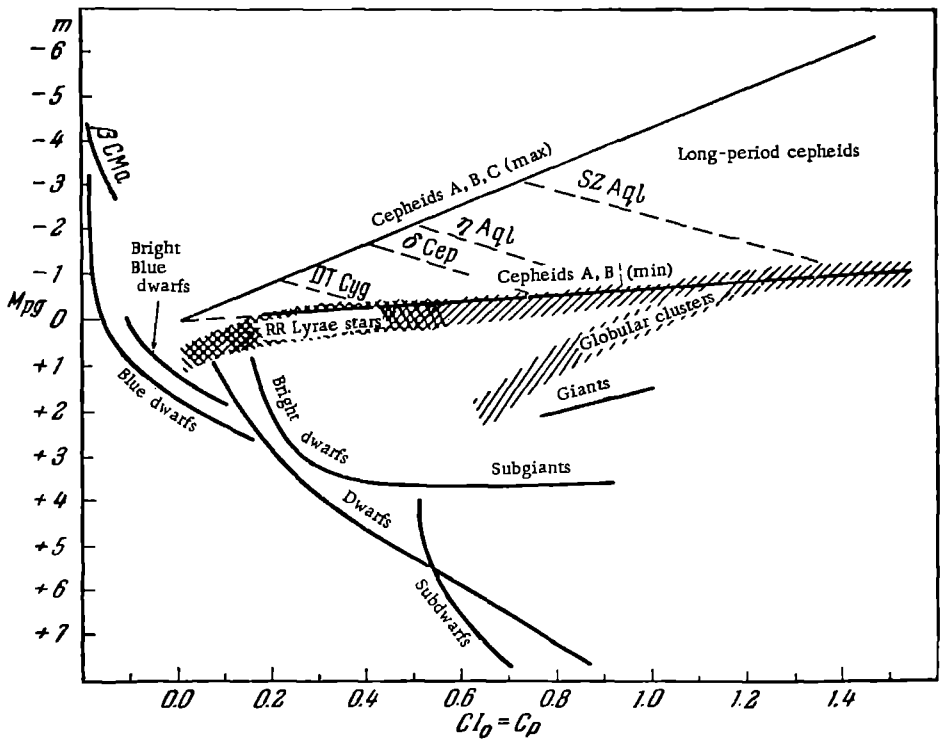


FIGURE 76

Other important characteristics of cepheids are those observed for pseudocepheids or supergiants. As an illustration Figure 77a shows the Hertzsprung—Russell diagrams for cepheids and pseudocepheids (c-stars). The spectra of pseudocepheids are close to cepheid spectra. Pseudocepheids include the supergiants listed in Table 18.

TABLE 18

Dwarfs			Subgiants			Giants			Supergiants		
Star	Sp MKK	C_p	Star	Sp MKK	C_p	Star	Sp MKK	C_p	Star	Sp MKK	C_p
Aver.	A2IV	0m.02	ε Cep	F0V	0m.22	δ Cyg	A0III	0m.08	41 Cyg	F4Ib	0m.33
Aver.	A4	0 .06	Aver.	F2	0 .28	γUMi	A3II—III	0 .04	α Per	F5Ib	0 .38
Aver.	F1V	0 .26	Aver.	F6IV	0 .40	α Oph	A5III	0 .10	45 Dra	F7Ib	0 .54
Aver.	F6V	0 .37	Aver.	F8IV	0 .43	γ Boo	A7III	0 .16	γ Cyg	F8Ib	0 .58
Aver.	G0V	0 .47	Aver.	G0IV	0 .52	Aver.	F2III	0 .24	β Aqr	G0Ib	0 .75
κ Cet	G5V	0 .58	Aver.	G5IV	0 .65	Aver.	F5III	0 .36	α Aqr	G2Ib	0 .89
Aver.	G8V	0 .64	β Aql	G8IV	0 .77	Aver.	F6III	0 .40	β Dra	G2Ib	0 .87
Aver.	K0V	0 .72	η Cep	K0IV	0 .81	ε Hya	G0III	0 .60	9 Peg	G5Ib	1 .04
ε Eri	K2V	0 .79	γ Cep	K1IV	0 .92	π Cep	G2III	0 .72	ζ Cep	K1Ib	1 .44
HR 8832	K3V	0 .88	—	—	—	η Her	G5III	0 .83	—	—	—
61 Cyg A	K5V	1 .05	—	—	—	Aver.	G8III	0 .84	—	—	—
						Aver.	K0III	0 .96			
						α Set	K3III	1 .16			
						Aver.	K5III	1 .40			

These are stars of luminosity class Ib in the MKK system. The same table also lists other stationary stars which can be used for converting Eggen's normal colors, i.e., C_p, to spectral types for various luminosity classes (1950).

Because of the period—spectrum and period—luminosity relations, cepheids of later spectral types are inevitably of higher luminosity.

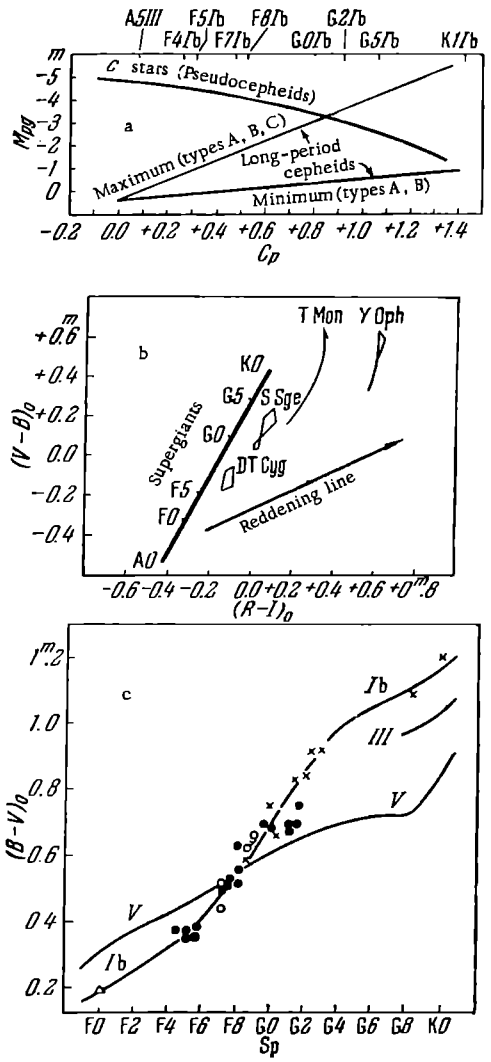


FIGURE 77

Pseudocepheids show a reverse dependence. The curves for the two groups of stars — stationary (or almost stationary, according to latest data) and nonstationary variables — cross in the region of G spectra, apparently a point of decisive significance in the theory of cepheid evolution.

Figure 77b further shows the position of the curves for some typical cepheids at different phases of the light curve in the $(V-B)_0$, $(R-I)_0$ diagram from the latest data of six-color photometry (Kron and Svolopoulos, 1959). The same figure shows the line of true colors of the normal supergiants (pseudocepheids) and the "reddening line" which accomodates the supergiants whose color is distorted by interstellar absorption (reddening). Note that the cepheid curves are stretched upward parallel to the supergiant line. We can thus find the color excess of the cepheids for any phase.

Figure 77c shows the true color $(B-V)_0$ and the MK spectra of a number of cepheids. The same figure marks the corresponding points for supergiants, giants, and dwarfs of luminosity classes Ib, III, and V. These data were borrowed from Kraft (1960) and obtained by narrow-band photometry.

We recall that the classical cepheids have the same distribution as B type stars relative to the plane of the Galaxy (and probably in space, although some contradictory data which require further explanation have been recently obtained). For this reason cepheids and B type stars are particularly convenient for

studying the interstellar absorption of light in the Galaxy. The cepheids give the total interstellar absorption (since their normal colors are known with insufficient accuracy) and the B stars can be used to determine the selective absorption (as their normal colors are known with higher accuracy). Numerous astrophysicists have worked on this problem.

Cepheids show prominent reddening, especially in Scutum, Cygnus, Sagittarius, Cassiopeia, and other regions of the sky. According to Badalyan (1956), the absorption for cepheids in the dark regions (Cygnus and Cassiopeia) is generally higher than in the luminous regions, even if the cepheids in the dark areas are closer to the Sun.

A highly interesting problem is to trace the spiral arms of the Galaxy using the distribution of cepheids.

The spiral arms have been originally charted using associations of hot O stars, supergiants, and early giants. The arms were also traced from some 150 diffuse nebulae, from star clusters, etc. There seems to be one outer arm and one inner arm a few hundred parsecs thick.

A considerable amount of information was obtained from radio observations of hydrogen clouds at 21 cm by Oort, van de Hulst, and Muller. These hydrogen clouds lie in the spiral arms of the Galaxy, and since the Doppler shift at centimeter wavelengths is very substantial, and the spiral arms move with different velocities, the clouds lying in different arms show through.

Studies of spiral structure were continued at various observatories both theoretically and experimentally. The results revealed the traces of five spiral arms.

In 1954 Torgor established that the long-period cepheids predominantly concentrated in the spiral arms.

A more detailed study of the concentration of cepheids in the spiral arms was carried out by Badalyan in 1956 (he used a new method). Working with the distance — galactic latitude diagram, he proved that cepheids like interstellar hydrogen clouds are concentrated in three outer and one inner spiral arm. The neutral hydrogen radiation intensity shows excellent correlation with the log period of the long-period cepheids. This, and other relations studied by Badalyan pointed to a possible genetic link between long-period cepheids and neutral hydrogen clouds. Recently, radial velocities and luminosities have been used to show that some cepheids apparently belong to galactic clusters, e.g., DL Cas belong to NGC 129. In 1960, Arp published his study on Cepheids in Galactic V Clusters, where three-color colorimetry with the 200-in telescope was applied to investigate in detail a long-period cepheid ($5^d.38$ period) and 40 nearby stars in the galactic cluster NGC 2355. It was apparently proved that the cepheid (CV Mon) belongs to this cluster.

9. The theory of cepheids. The pulsation theory has been developed in considerable detail. We will consider only the rudiments of this theory.

For small, radial and symmetric pulsations of a homogeneous gas sphere, where ρ is the mean density of the outermost layers, g is the gravitational acceleration, P is the gas pressure, we have at a distance R from the center

$$\frac{d^2R}{dt^2} = -g - \frac{1}{\rho}\frac{dP}{dR}.$$

It has been established theoretically that for low-mass stars, in particular for cepheids with masses of $5-8\,\mathfrak{M}_\odot$, the contribution of radiation pressure to the overall pressure is negligible.

The oscillations of the pulsating star are small and the physical parameters therefore can be expanded in a series about their equilibrium values (subscripted with 0).

Thus,

$$R = R_0(1+\alpha),$$
$$g = g_0(1+\alpha)^{-2} \approx g_0(1-2\alpha),$$
$$\varrho = \varrho_0(1+\alpha)^{-3} \approx \varrho_0(1-3\alpha).$$

Let the oscillations be adiabatic. Then $P \sim \varrho^\gamma$ (where γ is the specific heat ratio) so that

$$P = P_0(1+\alpha)^{-3\gamma} \approx P_0(1-3\alpha\gamma).$$

Now differentiating (for constant α), we have

$$dR = dR_0(1+\alpha),$$
$$dP = -g\varrho\,dR,$$
$$dP_0 = -g_0\varrho_0\,dR_0,$$
$$dP = dP_0(1-3\alpha\gamma) = -g_0\varrho_0\,dR_0(1-3\alpha\gamma).$$

Applying these transformations, we write the fundamental differential equation (omitting all terms with α^2) in the form

$$\frac{d^2R}{dt^2} = -g_0(1-2\alpha) + \frac{g_0\,dR_0}{dR}\frac{(1-3\alpha\gamma)}{(1-3\alpha)} = -g_0(1-2\alpha) + \frac{g_0(1-3\alpha\gamma)}{(1+\alpha)(1-3\alpha)} =$$

$$= -g_0(1-2\alpha) + \frac{g_0(1-3\alpha\gamma)}{(1-2\alpha)} = -g_0[(1-2\alpha) - (1+2\alpha-3\alpha\gamma)] = -g_0(3\gamma-4)\alpha = R_0\frac{d^2\alpha}{dt^2}$$

or alternatively

$$\frac{d^2\alpha}{dt^2} + \frac{g_0}{R_0}(3\gamma-4)\alpha = 0.$$

α characterizes the contraction of the cepheid and, like $\frac{d^2\alpha}{dt^2}$, it is independent of R. Therefore $\frac{g_0}{R_0}$ is also independent of R. But

$$\frac{g_0}{R_0} = G\frac{\mathfrak{M}}{R_0^3} = \frac{4}{3}\pi G\varrho_0 \qquad \left(\mathfrak{M} = \frac{4}{3}\pi\varrho_0 R_0^3\right).$$

If we assume a sinusoidal variation of α,

$$\alpha = \sin\omega t = \sin\frac{2\pi}{\Pi}t,$$

where Π is the period (to distinguish it from the pressure P). Insertion of α in the differential equation gives the oscillation period

$$\Pi = \sqrt{\frac{3\pi}{G\varrho_0(3\gamma-4)}}.$$

Hence we obtain the fundamental period—density relation:

$$\Pi \sqrt{\varrho_0} = \sqrt{\frac{3\pi}{G(3\gamma-4)}}.$$

If we take $\gamma = 5/3$ (a very low admixture of heavy elements), we find for $G = 6.66 \cdot 10^{-8}$

$$\Pi \sqrt{\varrho_0} = c_1 = 0.14.$$

For $\gamma = 14/9$ (a substantial content of heavy elements)

$$\Pi \sqrt{\varrho_0} = 0.24.$$

Alternatively, $\Pi \sqrt{\frac{\varrho_0}{\varrho_\odot}}$ is 0.12 and 0.20 in the two cases, respectively (here Π is expressed in days and $\varrho_\odot = 1.411\,\mathrm{g/cm^3}$).

We see that even the rudimentary theory of pulsation gives a constant c_1 which is close to the observed value. A more rigorous theoretical value for the standard stellar model using the radiation pressure according to Eddington (a "polytropic" sphere with $n = 3$ and $\varrho_c = 54.4\varrho$, where ϱ_c is the density at the center) is

$$\Pi \sqrt{\bar{\varrho}} = \frac{0.039}{\sqrt{3\Gamma-4}}$$

where

$$\Gamma = \beta + \frac{(4-3\varrho)^2(\gamma-1)}{\beta+12(\gamma-1)(1-\beta)}$$

(here β is a function of the radiation pressure).

According to later calculations by Severnyi ($\varrho_c = 54.18\bar{\varrho}$),

$$\Pi \sqrt{\bar{\varrho}} = \frac{0.0505}{\sqrt{3\Gamma-4}}.$$

Kozyrev's results (1948) show that the structure of cepheids is close to that of a polytropic gas sphere with $n = 3/2$, which corresponds to a slight concentration of stellar matter toward the center, $\varrho_c \approx 6\varrho$.

The pulsation theory is on the whole capable of explaining the characteristic features of cepheids. There are a number of substantial difficulties, however. These include 1) the relative phase shift in the variation of the photospheric and atmospheric radii and 2) the arbitrary pulsation amplitude (the equations are linear).

The first difficulty is explained by examining the behavior of a cepheid atmosphere in the variable radiation field. The arbitrary pulsation amplitude is more difficult to account for, since observations definitely show that the cepheid amplitudes, although spread between certain limits, are by no means arbitrary and are definitely a function of the period. Eddington's original pulsation equation allows pulsations of any amplitude about the state of stable equilibrium. Moreover, it predicts a considerable number of stars with nearly zero amplitudes among cepheids of any period. This contradiction between theory and observation shows that the cepheid pulsation is inherently associated with the instability of the star, i.e., the

equations describing the phenomenon are not linear differential equations with arbitrary amplitude. This was noted by Ambartsumyan back in 1930.

A substantial improvement of the pulsation theory was recently advanced by Zhevakin. He developed the theory of sustained oscillations of cepheids with a region of doubly ionized helium (assuming a generally high relative abundance of helium). He successfully explained the phase relations between the light curve and the radial velocity curve for the entire "large sequence" of variables. In our opinion, however, the pulsation theory should not be expected to do this, since the cepheid atmospheres exist under highly peculiar physical conditions and they require a special theory which does not relate their behavior to the properties of the interior layers, the interaction being highly complex.

The initial illusion of a perfect fit between the pulsation theory and the observations was shattered when discontinuous velocity variation was discovered for individual lines in the spectra of W Vir, RR Lyr, and other stars. There are reasons to believe that similar effects are also observed in the spectra of numerous other cepheids (the broadening and skewness of lines in their spectra suggests this); they have not been discovered so far only because of the insufficient spectrograph resolution: the line widths are too large, the exposures are too long (an inevitable shortcoming when photographing very faint objects), and so on. The last problem — finite resolution of spectral lines on the time scale (a most significant aspect of variables) — was studied by O. Struve (1954) for the particular case of BW Vul, a β CMa star.

The above effects cannot be explained by the pulsation theory in its pure and simple form. In our opinion, at least a qualitative explanation can be found if we further assume that powerful prominences are periodically ejected from the atmospheres of the pulsating cepheids, not unlike those which are frequently observed during the partial eclipse of 31, 32 Cyg, ε, ζ Aur, VV Cep and other stars. Another possibility is that the cepheids periodically shed their outer envelope, like certain Be stars. Magnetic fields apparently make a substantial contribution to the variation of cepheids (as to other stellar variation phenomena). Indeed, the motion of ionized stellar matter can be largely regulated (accelerated or decelerated) by magnetic fields. Small fluctuations of these magnetic fields will apparently explain the turbulence phenomena observed by O. Struve in various stars and by Mel'nikov in cepheids. A definite analogy can be drawn between this effect and the solar prominences (where the role of electromagnetic fields is obvious). Quite recently Babcock (1956) showed RR Lyr to possess a variable magnetic field, ranging from +1170 to −1580 ($\pm$ 150) gauss.

Note that the classical pulsation theory easily can be verified in observations (if the various difficulties mentioned in the preceding are ignored). A suitable criterion was proposed by Baade back in 1925. It is based on the fact that the variation of the radius derived from light and color curves on the one hand and by integration of the radial velocity curve on the other should give identical results (apart from possible phase differences). Bottlinger applied this criterion in 1928 to ζ Gem, though without success, since he associated this criteria with the assumption of blackbody radiation of the star. The radius and the line shifts varied out of phase. In 1940 Becker rejected the blackbody hypothesis and assumed only a

single-valued relation between the color temperature and the surface brightness. He obtained in-phase variation of the radius and the line shifts and actually calculated the radii in km.

As we have noted before, points of equal color index can be selected on the cepheid light curve (as Wesselink did for δ Cep). At these phases the surface brightness is presumably also equal, although the apparent magnitudes are different. The change in luminosity between the two corresponding points (phases) is thus attributable only to a change in the stellar surface:

$$\Delta m_R = -2.5\Delta \lg (4\pi R^2),$$
$$\Delta m_R = 2.17 \frac{\delta R}{R}.$$

Integration of the radial velocity curves gives the variation of the line shifts δD; hence approximately

$$\delta R = p\delta D \left(p = \frac{24}{17} \right) \quad \text{and} \quad \Delta m_R = 2.17 p \frac{\delta D}{\bar{R}} = 2.17 \frac{\delta D}{\frac{\bar{R}}{p}}.$$

Figure 78 plots Δm_R as a function of δD for δ Cep according to the last relation. We see that the dependence is indeed approximately linear and the ratio $\frac{\bar{R}}{p}$ (assumed unknown for δ Cep) is $18.8 \cdot 10^6$ km from the figure. If $p = 24/17$ we have $\bar{R} = 26.5 \cdot 10^6$ km for δ Cep.

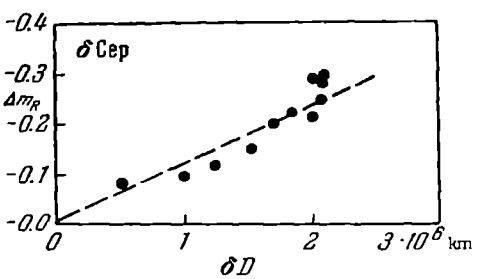


FIGURE 78

Thus, the pulsation theory fits the observations in this case.

An alternative pulsation criterion was proposed by M. Schwarzschild in 1938. He recommended using the radial velocity curve, and not the light curve, for control purposes. The period—density relation provides another highly attractive criterion of the validity of the pulsation theory. We have repeatedly mentioned various contradictions and inconsistencies. However, these contradictions are entirely eliminated if the zero point of the period—luminosity curve and its shape, as well as the shape of the mass—luminosity relation are appropriately adjusted in the calculation of $P\sqrt{\varrho}$.

Detailed comparison of the pulsation theory with observation data was carried out by Th. Walraven for δ Cep. He mainly compared the calculated effective kinematic value $g_k = G\frac{\mathfrak{M}}{R^2} + \ddot{R}$ with the value g_s obtained spectrophotometrically. The g_k of δ Cep (the second term $\ddot{R}$ dominates its variation) remains constant during most of the period, but just before the phase $0^d.9$, i.e., before the light maximum, where the radius obtained by integration of the velocity curve is at its minimum, it abruptly and markedly increases. The variation of g_s follows a much smoother curve, however, and its maximum is delayed: it appears simultaneously with the light

maximum or even later, when the radius is close to its mean value. Moreover, g_s is on the average 1/40 of the previous value. Thus, unlike Kipper (1934) or Krüger (1937), who used an inaccurate relation ($W \sim \sqrt{N}$) for converting the line widths or equivalent widths to number of atoms, Walraven did not confirm the hypothesis of the direct relation between the effective g_s and the acceleration. His results thus do not support the hypothesis of simple pulsations either. The correctness of his findings is apparently supported by O. Struve's result (1944), who noticed a substantial enhancement of the ionized lines just before the light maximum (when P is smallest) in 7 cepheids, and also by the result of van Albada (based on Pannekoek's data), who also established that for δ Cep near the minimum radius the g_s was less than soon after the maximum. According to Walraven, these data point to excellent agreement with Schwarzschild's pulsation model, i.e., the model of a travelling (surfacing) wave, as is evident from the similarity in the curves of electron pressure and $\dot{R}$. The marked increase in P_l from minimum to maximum corresponds to a sharp increase in $\dot{R}$ and g_s. On the other hand, Schwarzschild's theory has its own unsolved difficulties and contradictions.

The recent development by Frank-Kamenetskii (1955) of the theory of "central pulsations" opens new possibilities toward ensuring a better fit between theory and observations. In particular, the theoretical and the observational radial velocity curves show a good fit. An attempt was made to extend this theory to the case of radial velocity curves of cepheids with secondary maxima (Iroshnikov, 1959).

The pulsation theory definitely explained some fundamental features of cepheids. The great variety of observational material, however, requires further radical improvement of the existing theory or possibly its replacement by a better theory.

10. Characteristics of cepheids in other star systems. This extensive subject will be considered very briefly.

Short-period and long-period cepheids were observed in galactic globular clusters and also in globular clusters in extragalactic star systems. Long-period cepheids were also observed in the nearest extragalactic neighbors, the Ir star systems in the Large and the Small Magellanic Clouds, and also in NGC 6822 and NGC 1613. Fairly recently, globular clusters were observed to contain occasional long-period cepheids; long-period cepheids were also found in the nearest spiral galaxies, M 31 in Andromeda (Sb), M 33 in Triangulum (Sc), M 81 in Ursa Major (Sb), and others.

These cepheids on the whole show the general characteristics described for the galactic cepheids; in particular they obey the period—luminosity relation. Deviations of the cepheids in Magellanic Clouds and in the spiral galaxies in Andromeda and Triangulum from this relation in fact enabled Gordon (1945) to devise a new technique for confirming the presence of light absorbing matter in these systems.

a) CEPHEIDS IN GLOBULAR CLUSTERS. Short-period cepheids were discovered in great numbers in globular clusters. According to Sawyer's catalogue (1955), 72 globular clusters contain 1421 known variables. Seven clusters contain over 50 variables. The clusters richest in variables are M3 (NGC 5272) in Coma Berenices (187 variables), NGC 5139 or ω Centauri (164 variables), M5 (NGC 5904) in Serpens (97 variables), and M15 (NGC 7078) in Pegasus (93 variables).

Some 10% of the cluster variables cannot be regarded as short-period cepheids. In particular, 36 clusters contain 122 stars of this exceptional group. Of these 15 have periods greater than 10 days, and 13 periods from 10 to 26 days. They are classified as long-period cepheids or W Vir variables of the halo component.

The distribution of all the known short-period cepheids in globular clusters according to Sawyer's data is shown in Figure 79.

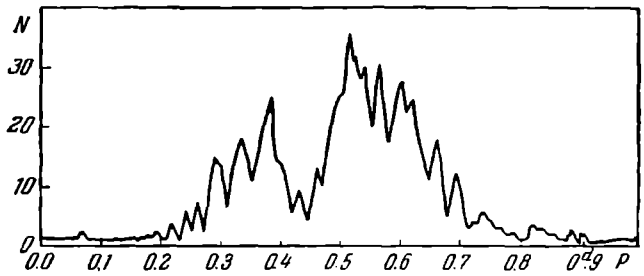


FIGURE 79

The most remarkable feature is the large minimum near periods of $0^d.44$; it intervenes between the two maxima at $0^d.39$ and $0^d.52$. This bimodal distribution indicates that the short-period cepheids do not constitute a homogeneous population, as the minimum cannot be attributed to selection effects (Kukarkin, Struve, Joy, Shapley, and others). Smith (1955) has shown, in particular, that galactic cepheids with periods less than 0 .2 are not RR Lyrae stars. Their period—spectrum relation is similar to that of RR Lyrae stars (types a and c), but it is displaced along the spectrum axis.

The period—color relation for stars with $P < 0^d.2$ has a smaller slope than that for RR Lyrae stars, which in its turn is close to that for M3. The light variation amplitudes of cepheids with $P < 0^d.2$ decrease from 1 to 0.3 stellar magnitudes as the period is increased between $P = 0^d.05$ and $0^d.2$.

Analysis of the motion of these cepheids (unfortunately, only 9 stars) gave 66 km/sec for the velocity of the Sun. The corresponding apex was found to lie between the apices for the members of the Local Group and for the high-velocity stars. The proper motions reduced to one stellar magnitude decrease with increasing period, a point of considerable significance in the application of the average parallaxes. Their radii, obtained by the method of Baade—Wesselink—Stebbins (see above), are 1.4 and 3.8 $R_\odot$ for CY Aqr and VZ Cnc, respectively. The absolute magnitudes of stars with $P < 0^d.2$ are generally between $+1^m$ and $+4^m$. Stars of this group are very seldom observed in globular clusters. This is an entirely special group of stars: they are small, with intermediate masses. Their variation, except the length of the period, is reminiscent of the variation of RR Lyrae stars, but in terms of their galactic distribution, motion, and masses they are closer to the classical long-period cepheids. Smith appropriately proposed the name of "dwarf cepheids" for these stars (typical representatives CY Aqr and AI Vul). Indeed, we know that the c characteristic is

observed only for long-period cepheids with medium periods; short-period cepheids and classical long-period cepheids (those with long periods) do not possess this characteristic.

The entire frequency curve of Figure 79 clearly can be split into sections differing in certain characteristics. Note that the frequency curves of short-period cepheids in individual clusters vary between wide limits. These differences are correlated only with the shape of the color—luminosity diagram of the particular cluster. According to Oosterhoff (1939), Sawyer (1944) and others, a number of clusters show a unimodal distribution of log periods; these are mainly clusters rich in variables (M3, M5, NGC 3201, etc.). Unimodal distributions are classified as Ia by Sawyer. Other clusters, such as ω Cen and M15, have bimodal distributions of log period. These distributions are classified as Ib (a small secondary maximum) and II (two comparable maxima). The different frequency curves are shown in Figure 80 for three particular clusters. For comparison Figure 81 shows the distribution of the galactic cepheids near the Sun and in the nucleus of the Galaxy (no cepheids are observed at the very center because of extremely high absorption by dark clouds).

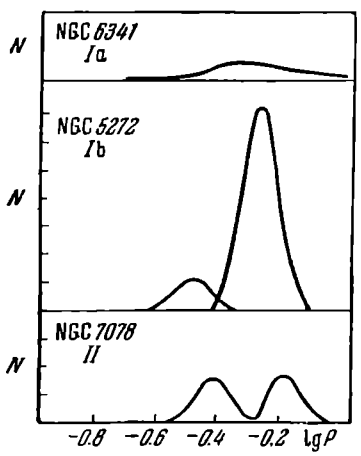


FIGURE 80

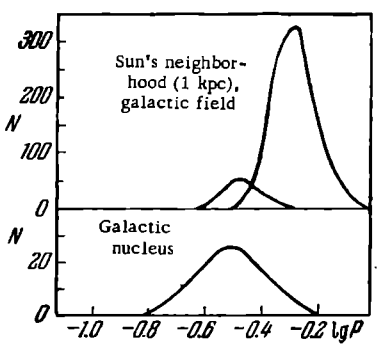


FIGURE 81

We see that the frequency curve of the nucleus cepheids is clearly of type Ia, and that of the galactic field cepheids is of type Ib, although the two maxima are somewhat displaced. These data, as well as many other interesting facts, will be found in the recent monograph by C. Payne—Gaposchkin. The frequency maxima correspond to the following values of $\lg P$: −0.25 (Ia), −0.28 and −0.58 (group Ib), −0.21 and −0.54 (group II), −0.27 and −0.47 (the galactic field), and −0.50 (the galactic nucleus). Some frequency data for a number of globular clusters of different types are given in Table 19.

Hertzsprung's relation for the shapes of the light curves of RR Lyrae stars in globular clusters is approximately the same as for the free (unbound) galactic cepheids.

TABLE 19

Type	NGC	Period corresponding to maximum frequency	Ellipticity	Concentration	Spectrum	Absolute magnitude
Ia	362	$0^d.50$—$0^d.55$	8	III	—	$-7^m.3$
Ia	3201	0 .50—0 .55	9	X	—	—
Ia	6341	0 .50—0 .55	8	IV	A9, A5n	— 7.8
Ia	6723	0 .50—0 .55	9.5	VII	G3	—
Ia	7089	0 .60—0 .65	9	II	F3, dF0	— 8.5
Ib	5272	$0^d.50$—$0^d.55$; $0^d.30$—$0^d.35$	8	VI	F5, dF2	— 8.2
Ib	6121	0 .50—0 .55; 0 .25—0 .30	9	IX	—	—
Ib	5904	0 .45—0 .50; 0 .30—0 .35	9	V	F6, dF7	— 8.0
Ib	6981	0 .50—0 .55; 0 .35—0 .40	—	II	G2	— 6.6
II	5053	0 .55—0 .75; 0 .25—0 .45	8	VI	—	— 5.3
II	5466	0 .60—0 .65; 0 .30—0 .35	8	VII	F6	—
II	5024	0 .60—0 .65; 0 .35—0 .40	9	V	F8, A2n	— 7.8
II	5139	0 60—0 .65; 0 .35—0 .40	8	VIII	—	—10.0
II	7078	0 .65—0 .70; 0 .35—0 .40	8	IV	F0, dF0	— 8.3

In the globular cluster ω Cen the mean skewness is $\varepsilon = \frac{t_M - t_m}{P} = 0.47$ for type *c* stars and 0.12 for type *a* stars. In the galactic field, these values are 0.455 and 0.22, respectively. The variables in the galactic nucleus have skewed curves and thus do not belong to type *c*. Note that M. Schwarzschild showed in 1940 that type *c* stars in the globular cluster M3 are bluer than the type *a* stars in the same cluster.

The star with the shortest period in ω Cen is N65 ($P = 0^d.0627$); its light curve, like the light curves of all other stars with exceedingly short periods in globular clusters, is symmetric, i.e., of type *c*. RR Lyrae stars with longer periods have a high skewness (type *a*), which, however, diminishes again at still longer periods (type *b*). The boundaries between types *a*, *b*, and *c* differ from cluster to cluster together with the frequency characteristics, the cepheid count, the magnitude—color diagram, and other properties. The division of cluster variables into types *a*, *b*, and *c* is much more pronounced than for the galactic cepheids.

The spectra of RR Lyrae stars in globular clusters are apparently similar on the whole to the spectra of free galactic stars. This is evident from Table 20 (A. Joy's data for galactic cepheids) borrowed from Payne—Gaposchkin's book. We see that estimates based on hydrogen lines, metallic lines, and H and K lines give different results.

TABLE 20

Period	Spectrum from hydrogen lines		Spectrum from metallic lines		Spectrum from H and K lines	
	maximum	minimum	maximum	minimum	maximum	minimum
$0^d.20$—$0^d.39$	A4.0	F3.5	A8.5	F2.0	A4.0	A7.5
0 .40—0 .59	A5.5	F6.0	A8.0	F3.5	A2.0	F 0.0
0 .60—0 .80	A5.5	F3.0	A9.0	F2.5	A3.5	F 2.0
Average	A4.7	F4.2	A8.5	F2.8	A3.2	A9.2

The RR Lyrae stars in globular clusters are much too weak to permit taking their spectrograms, and their color is thus the only source of our information on spectra. The globular clusters also contain Population II long-period cepheids, which are brighter and thus could be studied spectroscopically (Joy, 1949). Joy divided all the "pulsating" cluster variables into 4 groups: the shortest period group of the long-period cepheids, $P < 5^d$, W Vir stars with $13^d < P < 19^d$, RV Tau stars with $25^d < P < 90^d$, and semiregular variables with $65^d < P < 106^d$. The averaged data are listed in Table 21 (27 stars).

TABLE 21

Period (days)	Type	Average period $\bar{P}$	Spectrum amplitude Δ Sp	Average spectrum	Absolute photographic magnitude *	Number of stars N
1—2	Short-period	$1^d.82$	A7.2—F2.5	A9.8	$-0^m.9$	6
13—19	W Vir	16 .35	F7.2—G2.2	F9.7	−1 .9	10
25—90	RV Tau	53 .60	F5.2—G4.0	F9.6	−3 .0	5
65—106	Semiregular	88 .50	G2.0—G6.2	G4.1	−1 .5	6

* In relation to RR Lyrae stars in the same cluster.

Population II long-period cepheids in globular clusters are different from the Population I long-period cepheids of the disk subsystem. They are reminiscent of W Vir stars, i.e., long-period cepheids of the halo component. Their spectra (7 out of 10), like those of galactic W Vir stars, show bright hydrogen lines on the ascending branch of the light curve.

The curves of long-period cluster cepheids do not follow the Hertzsprung—Ludendorff relation (more than 50 such stars with $P > 1^d$ are known at present), i.e., the shape of their curve is not a function of the period. The number ratio of these stars to RR Lyrae stars in globular clusters is higher than in the Galaxy. Their spectra are earlier than the spectra of galactic cepheids of the same period, especially at the minimum, and the G band (CH) is not observed. Bright lines in galactic stars of this type are observed in the spectra of W Vir ($17^d.27$), RX Lib ($24^d.95$), MZ Cyg ($27^d.17$), V377 Sgr ($16^d.17$), AA Ser ($17^d.16$), and others. Their light curves are similar to the light curves of long-period cepheids in globular clusters. The luminosity of the cluster cepheids can be accurately determined in relation to that of the short-period cepheids of the same cluster. As lg P varies from 0.0 to 2.0, their luminosity relative to RR Lyrae stars varies from $0^m.0$ to $-3^m.0$ or even $-4^m.0$. W Vir stars thus obviously follow the period—luminosity relation, which is probably the same as that for RR Lyrae stars. It is moreover a single-valued relation, and not a triple-valued sequence as suggested by Arp in 1955 (see above). Their luminosity, however, is weaker by $0^m.5$ (Mel'nikov) or by $1^m.5$ (Baade) than the luminosity of the long-period cepheids of the disk component.

Numerous Population II cepheids with $P > 15^d$ have double periods (Arp, 1955) with marked variation of even and odd epochs. They, like W Vir, have broad light maxima. The light curves are not as skewed as the curves of Population I cepheids. They are sometimes "bimodal," as for one of the stars in M12. Population II stars in globular clusters will suddenly flare

up (e.g., stars with $P = 5^d.1$ in M13) during 5 hours. This phenomenon was first observed for the galactic cepheids by Hogg in 1929 and then by Bernheimer for η Aql (1931). The light curves of Population II cepheids in globular clusters show a delay in the long-wave spectral region. Here the delay between $\lambda = 4250$ Å and $\lambda = 5500$ Å, according to Arp, is greater than that for Population I long-period cepheids (δ Cep and η Aql, $0^d.05$) between 3530 and 10,300 Å. Some authors are of the opinion that a gradual transition is observed from W Vir stars to RV Tau stars. Many others, however, disagree on this point.

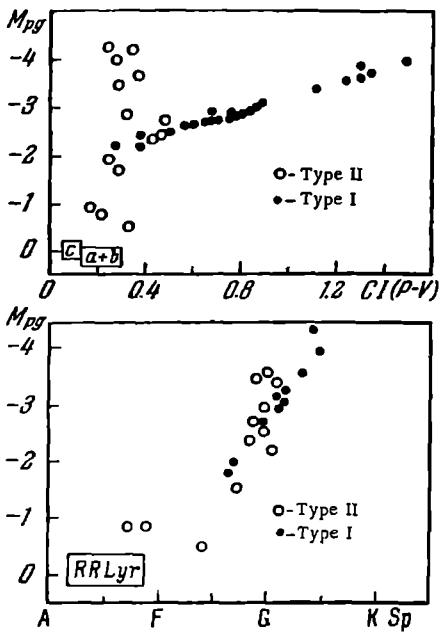


FIGURE 82

Figure 82 gives the color—luminosity and the spectrum—luminosity diagrams for Population II cepheids (the halo component) in globular clusters and for Population I cepheids (the disk component), according to Arp's latest data (1955).

b) CEPHEIDS IN MAGELLANIC CLOUDS. These irregular star systems, joined to one another and possibly to the Galaxy by diffuse clouds of matter (radio observations), contain up to 1000 variables of various types each. Most of these variables are long-period cepheids. However, weaker and therefore less accessible short-period cepheids were also photographed in 1952—1953 by Thackeray at the Radcliffe Observatory, first in globular clusters associated with the Small Cloud (NGC 121), and then in the Large Magellanic Cloud (NGC 1466 and NGC 1978). These are halo-component stars, but the Small Magellanic Cloud also contains short-period cepheids with periods of 0.647, 0.747, and $0^d.913$ (Dartayet and Dessy, 1952), which are apparently stars of the disk component, since they are $1^m.6$ brighter than the corresponding normal RR Lyrae stars. The normal RR Lyrae stars number 3, 21, and 2 in NGC 121, NGC 1466, and NGC 1978, respectively. A frequency curve of the cepheid periods in the Small and Large Magellanic Clouds and in the Galaxy is shown in Figure 83. We see that the curves are entirely different. The small Magellanic Cloud has an excess of long-period cepheids with shorter periods compared to the cepheids of the Large Magellanic Cloud. The frequency curves of the two Clouds (maxima at $2^d.5$ and $3^d.8$ for the Small and the Large Magellanic Clouds, respectively) are different from the galactic curve (maximum at $4^d.2$); they further change when moving from the center of the Cloud to the periphery (the curves shift in the short-period direction). The light curves of the long-period cepheids in Magellanic Clouds are on the whole similar to the light curves of the galactic cepheids and they are mainly classified as Population I, i.e., stars of the disk component, or Population II, i.e., stars of the halo components.

The periods of the long-period cepheids in the Magellanic Clouds range from slightly longer than 1^d to $118^d.64$, whereas the longest known period in the Galaxy is only $45^d.1$ (SV Vul). That these are cepheids is evident from their exceedingly high luminosity.

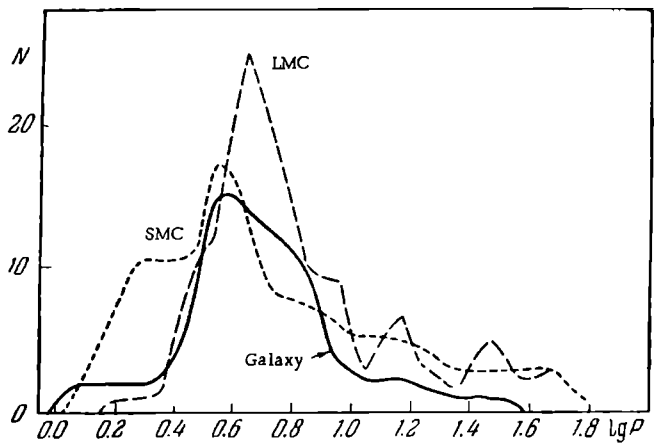


FIGURE 83

No spectra of cepheids in the Magellanic Clouds have been obtained due to the extreme faintness of the stars and the absence in the Southern Hemisphere of telescopes comparable with the 200-in Palomar reflector. The colors of the cepheids in the Magellanic Clouds were studied by a long line of astronomers. Highly interesting results were obtained in 1959 by Kron and Svolopoulos, who used six-color photometry. They showed (Figure 84) that on the $P - V$ vs. lg P diagram the points corresponding to the cepheids in the Small Magellanic Cloud (light circles) lie below the corresponding dark circles of the galactic cepheids at the maximum and the minimum (9 and 25 stars, respectively). This signifies that the cepheids in the Small Magellanic Clouds are much bluer than the galactic cepheids. The colors of the cepheids in the Small Magellanic Cloud were corrected for interstellar absorption ($0^m.09$). The only exception to this rule are Eggen's type C galactic cepheids — AL Vir and Y Oph. The difference in color for the cepheids in the Galaxy and in the Small Magellanic Cloud is not the same at the maximum and the minimum. Hence the conclusion that

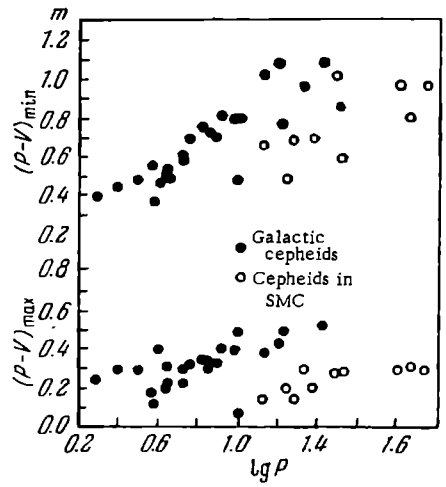


FIGURE 84

this difference cannot be attributed to improper correction for the selective interstellar absorption. Even the most extreme assumptions are incapable of accounting for the observed difference in color, contrary to the belief of numerous older astrophysicists. The cepheids in the Small Magellanic Cloud are thus significantly bluer than the long-period cepheids in the Galaxy, which is apparently due to differences of age.

c) CEPHEIDS IN EXTRAGALACTIC NEBULAE. No short-period cepheids have been discovered so far in other spiral galaxies because of their extreme faintness. They were found, however, in low-luminosity elliptic systems in Sculptor (type *a* and *c* RR Lyrae stars). Their periods are the same as in M3, but the frequency curve is different from that of the cepheids in ω Cen.

Long-period cepheids of the disk component (Population I) were discovered in great numbers in such spiral galaxies as M31, M33, M81, and others, since they are of much higher luminosity than their short-period counterparts. In Andromeda Nebula (M31) up to 38 Population I cepheids were observed. In Triangulum (M33) up to 50 variables were discovered, mainly also long-period cepheids. The periods of the cepheids in these and similar systems range from a few days to values much longer than the periods of the cepheids in the Magellanic Clouds ($118^d.64$). M31 contains a star with $P = 175^d$ ($m = 17^m.9$ at the maximum). The long-period cepheids in spiral nebulae follow the period—luminosity relation.

The light curves of cepheids in spiral nebulae are not unlike the light curves of the Population I galactic cepheids (the disk component). Numerical comparison shows that the long-period cepheids in M31 and in galactic globular clusters belong to different period—luminosity relations (specifically, the disk and the halo components, Populations I and II).

11. Frequency distribution of the cepheids of the large sequence. β CMa stars, RR Lyr stars, long-period cepheids, W Vir stars, RV Tau and SX Her semiregular variables, red semiregular variables, and long-period variables are often referred to as "pulsating variables" (Payne-Gaposchkin and others). The frequency curve of their periods is schematically shown in Figure 85 (after Payne—Gaposchkin). To avoid selection associated with the effect of the absolute magnitude, all the figures were reduced to a common volume in the Galaxy. The effect of this correction is evident from the following. The maximum of the galactic cepheids corresponds to $\lg P = 0.72$; after reduction to a common volume it shifts to $\lg P = 0.64$. This is so because RR Lyrae stars have lower absolute luminosities than the long-period cepheids, and their relative number is therefore smaller.

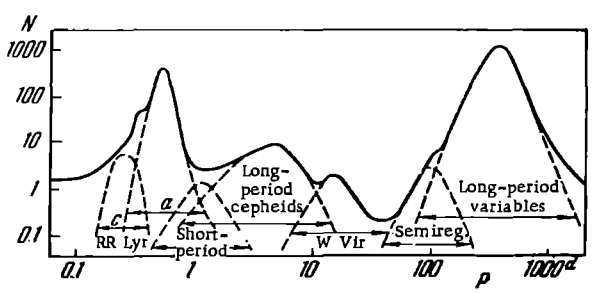


FIGURE 85

Frequency and distribution studies must always take into consideration other distorting effects, e.g., detection probability, etc. Unfortunately, the corresponding reductions are not always introduced with adequate accuracy and certainty.

REFERENCES TO CHAPTER VII

1. BADALYAN, G.S.— Bull. Byurakan Observ., No.8:27. 1951; No.17:3, 27. 1956.
2. GORDON, I.M.— Bull. Kharkov Astron. Observ., 7:10. 1947.
3. ZHEVAKIN, S.A.— Astron. Zh., 30:161. 1953; 31:335. 1954.
4. IVANOVSKAYA, V.I.— Trudy IV Soveshchaniya po Voprosam Kosmogonii, 304. 1955.
5. KUKARKIN, B.V.— Astron. Zh., 13:111. 1936; 14:125. 1937; Peremennye Zvezdy, 7:57, 69. 1949.
6. KUKARKIN, V.V. Issledovanie stroeniya i razvitiya zvezdnykh sistem na osnove izucheniya peremennykh zvezd (Structure and Evolution of Star Systems from a Study of Variables).— *Gostekhizdat.* 1949.
7. KUKARKIN, B.V. and P.P. PARENAGO. Peremennye Zvezdy (Variable Stars), Vol. 1.—ONTI. 1937.
8. LEBEDINSKII, A.I. and L.E. GUREVICH.— Astron. Zh., 26:97, 138. 1949.
9. MEL'NIKOV, O.A.— Trudy, Pulkovo Observ., 64. 1950; Trudy IV Soveshchaniya po Voprosam Kosmogonii, 296. 1955.
10. OGORODNIKOV, K.F.— PAAS, 7:105. 1932 (in collaboration with O. Struve); Astron. Zh., 10:304. 1933.
11. PARENAGO, P.P.— Peremennye Zvezdy, 10:193. 1955.
12. PSKOVSKII, Yu.P.— Astron. Zh., 36:662. 1959.
13. ROSSELAND, S. The Pulsation Theory of Variable Stars.— Oxford University Press. 1949.
14. SEVERNYI, A.B.— Izvestiya, Crimea Astrophys. Observ., Part I, 2. 1948; Astron. Zh., 23:11. 1946.
15. FRANK-KAMENETSKII, D.A.— Doklady AN SSSR, 86:897. 1950; 99:41, 222. 1954.
16. ALLER, L.H.— Astrophysics, Vol.2. 1954.
17. BAADE, W.— PASP 68:1. 1956.
18. CAMPBELL, L. and L. JACCHIA. The Story of Variable Stars.—Harv. Mon. 1945.
19. CANAVAGGIA, R. and J.C. PECKER.— Ann. d'Astrophys, 15:260. 1952; 18:431. 1955.
20. JOY, A.H.— ApJ. 111:105. 1949.
21. IVANOVSKA, W.I.— Tr. IAU, 8:814. 1952.
22. PAYNE-GAPOSCHKIN, C. and S. GAPOSCHKIN. Variable Stars.—Harv. Mon. 5, 1938.
23. REDDISH, V.C.— MN 115:480. 1955.
24. SAWYER, H.B.— Journ. RAS Can. 49:114. 1955.
25. SCHWARZSCHILD, M. and B., W.S. ADAMS.— ApJ, 108:207. 1948.
26. STRUVE, O.L.— Tr. IAU, 8:808, 1952; Sky and Telescope, 12, 203, 23. 1952.
27. WALRAVEN, Th.— Publ. Amst. Inst. 8, 1948.
28. WESSELINK, A.J.— BAN, 10:252. 1947; 10:91, 1946.
29. PRESTON, G.W.— ApJ, 130:507. 1959.
30. KRON, G.E. and S.N. SVOLOPOULOS.— PASP, 71:126. 1959.
31. KRAFT, R.P.— ApJ, 131:330. 1960.

V.G. GORBATSKII

Chapter VIII

LONG-PERIOD, SEMIREGULAR, AND IRREGULAR HIGH-LUMINOSITY VARIABLES

§47. Long-period variables. Long-period variables were the first among variable stars to attract the attention of astronomers. This is understandable, since the large light-variation amplitudes make their detection a comparatively simple undertaking. Some 3700 long-period variables had been discovered by 1959. They account for 25% of all the listed variables.

A variable star is classified as a long-period variable if it satisfies the following conditions:

a) the light variation period is over 70 days;

b) the spectrum shows bright hydrogen lines near the maximum.

The period of the long-period variables is not constant: it fluctuates about the mean value (with deviations reaching up to 10% of the mean). The longest known period is 1379 days, whereas most of the periods lie between 200 and 400 days.

The absorption spectra of the long-period variables fall into three groups:

1. Me stars, whose spectra show strong TiO absorption bands.

2. Re and Ne stars, whose spectra show molecular absorption bands of C_2, CN, and CH; Re and Ne stars recently have been combined into a single spectral type, Ce.

3. Se stars, with strong ZrO absorption bands.

About 90% of all the known long-period variables are Me stars. The light variation amplitudes and the shape of the light curves for Me and Se stars are not markedly different. The light variation amplitudes of Ce stars are somewhat smaller than those of Me stars. The spectroscopic characteristics, and in particular the relative intensities of the bright lines, however, are different in the spectra of Me, Se, and Ce stars. The main features of the long-period variables of each type are considered in what follows.

Light curves are a highly important tool in the study of long-period variables, and of any other group of variables for that matter. Figure 86 is a specimen light curve of the variable X Cam; the shape clearly changes from one cycle to the next.

FIGURE 86

Similar changes in the shape of the light curve are observed almost for all the long-period variables. Since the shape of the curve is variable and the period also changes, we are generally dealing with average light curves of these stars. Figure 87 shows the average light curves of S UMa (a) with $P = 225$ days and χ Cyg (b) with $P = 407$ days.

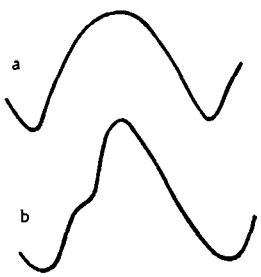


FIGURE 87

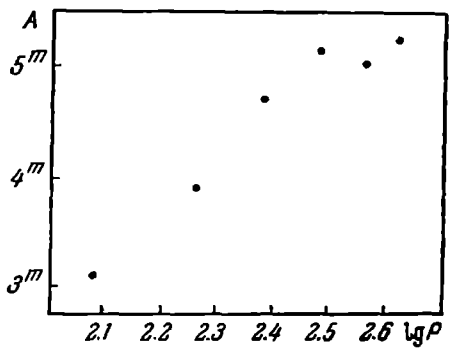


FIGURE 88

Despite the great number of statistical studies of the average light curves and their classification, the physical meaning of the principal features of these curves remains still unknown, although it is clear that many of these features are closely related to the actual variation mechanism. These include the increasing skewness of the light curve with the increase of the period (the rise branch becomes steeper than the descending branch) and the presence of a hump on the rise branch for almost all stars with periods exceeding 400 days. The dependence between the mean variation amplitude and the average period is also of some interest. This dependence is shown in Table 1, which lists the average amplitudes A for groups of stars with different periods P (n is the number of stars in each group). Figure 88 is the plot of these data.

TABLE 1

P	$\bar{A}$	n
91—150	3.1	12
151—210	3.9	39
211—270	4.7	55
271—330	5.1	63
331—390	5.0	48
391—450	5.2	29
451—570	5.2	11

Alongside with light curves in the visual light, there are light curves in other spectral regions. The difference between the light curves in photographic and visual light is not large, and the photographic amplitudes are close to the visual. On the other hand, the light variation of the long-period variables in the infrared is much less pronounced than in the visual spectrum /2/. The mean amplitude at effective wavelength $\lambda = 8500$ Å is half the mean visual amplitude ($2^m.4$). At $\lambda_{eff} = 9500$ Å, the light variation amplitude of the three stars was found to be a mere $1^m.4$.

The brightest long-period variables were also studied with radiometers. The "radiometric magnitude" curves were plotted for 11 stars /3/, and the "radiometric amplitude" of these stars was found to be about 1^m.

As we know, a radiometer responds to the total radiation transmitted by the Earth's atmosphere, i.e., the integrated light of the star, except its

ultraviolet radiation and radiation in parts of the infrared region. The likeness of "radiometric" and infrared amplitdes indicates that the bulk of the stellar energy reaching the Earth is radiated in the far infrared (10,000—15,000 Å). This is not an unexpected conclusion, since the red color of the long-period variables immediately suggests that their intensity peak lies in the infrared region.

Since the energy radiated by a star in the visual spectrum is only a minor fraction of the total radiation of long-period variables, the visual magnitude of the star may change considerably even though the total radiation fluctuates between small limits. The visual amplitudes calculated from the measured radiometric amplitudes assuming a Planck law radiation were indeed found to be much greater than the radiometric amplitudes, but still smaller (by about 2 m) than the observed visual amplitudes.

Some astronomers, in particular Merrill /5/, are of the opinion that the divergence between the calculated and the observed visual amplitudes is associated with the decrease in the transparency of the stellar atmosphere in the visual spectrum as it cools down. Some attribute the increasing opacity to absorption by TiO molecules, whereas others suggest that the radiation is absorbed in clouds of solid or liquid particles. These factors should cause additional redistribution of the radiation energy between the different spectral regions, definitely reducing the intensity in the visual spectrum. So far, these assumptions are only working hypotheses, since there are neither theoretical nor observational findings to support them.

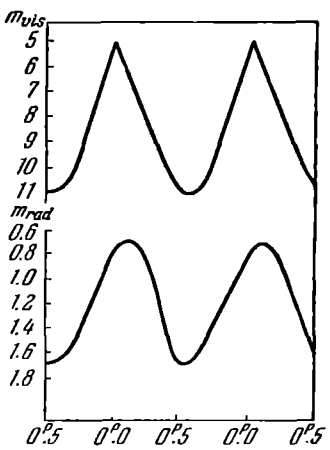


FIGURE 89

The variation curves of the visual and the radiometric magnitude (average for 11 stars) shown in Figure 89 differ in shape, as well as in amplitude. The radiometric peak markedly lags behind the visual peak.

The radiometric magnitudes were used to calculate the bolometric magnitudes of the stars and the corresponding light curves were plotted assuming negligible radiation in the far ultraviolet. The bolometric curves, as was expected, are very much like the radiometric curves.

To judge from the average light curves of long-period variables, their magnitude varies slowly and monotonically. In some cases, however, discontinuous jumps were observed. For instance, the rise branch of UX Cyg in one of the cycles showed a drop of 1m.4 in 12 minutes; examination of old photographs confirmed the previous occurrence of this effect /4/. A marked increase in brightness — sixfold in two days — was observed on the rise branch of R Vir in one of the cycles.

We do not know whether or not most stars of this type are characterized by such rapid changes of magnitude. These jumps apparently occur mostly near the minimum or on the rise branch (the change in radiation required to produce a change of 1—2 magnitudes near the minimum would pass almost unnoticed near the maximum). In these phases the stars are weak

and are not generally covered by amateur observers. Photoelectric measurements of these stars at and near the minimum will provide a wealth of valuable information on light variation of the long-period variables.

We will now proceed with a discussion of the spectra of long-period variables. We will only concentrate on the basic facts, since the observational data are most extensive. A comparatively comprehensive description of the spectra of long-period variables will be found in Merrill's monograph /5/ published in 1940, and in a number of later studies on the spectra of individual long-period variables.

Spectrophotometry of the continuous spectrum of long-period variables and especially of Me stars involves considerable difficulties because of the large number of absorption lines and molecular absorption bands in the visible spectrum ($\lambda\lambda$ 4600—6500 Å). No exact determinations of the color temperature of long-period variables are therefore available, but judging from the red color of these stars, it should range around 2500—3000°. The long-period variables grow redder along the descending branch, and the molecular absorption bands become stronger. In terms of TiO absorption bands at the minimum, the long-period variables can be classified as M5e to M9e.

Observations show that the intensity of absorption bands of a given star averaged over the period (and hence the mean spectral subtype of the star) depends on the length of the period: the longer the period the higher is the (average) intensity of the absorption bands. This dependence possibly accounts for the previously mentioned relation of the light variation amplitude to the period. Indeed, if the luminosity of two stars of different subtypes is decreased by the same amount, the intensity of the absorption bands will increase to a greater extent in the star of the later subtype; therefore if the observed light variation is partly attributed to variation of absorption in the molecular bands, the star of the later subtype should reveal a larger light variation amplitude.

The continuous spectrum of Se long-period variables has a most remarkable property: the continuum intensity shortward of λ 4600 Å falls off very rapidly, and for $\lambda \ll 4000$ Å it is vanishingly small. No such effect is observed in the spectra of Me stars, and their relative continuum intensity between λ 4200 Å and λ 4000 Å (largely free from TiO absorption bands) is comparatively high; it even increases at the minimum. A similar effect of a rapid reduction in the continuum intensity of normal N stars has been recently explained by absorption of radiation in molecular carbon compounds (C_3, CH). It seems that the atmospheres of Se stars contain some other, hitherto unknown, absorption mechanism, which is particularly effective at $\lambda < 4000$ Å.

The absorption spectrum of long-period variables, besides molecular bands, contains a large number of absorption lines, mainly of neutral and partly ionized metals. Iron lines are particularly numerous. The absorption line spectrum reaches its peak development just before the maximum, when it is essentially similar to the absorption spectrum of normal red giants of the corresponding spectral subtype.

In the postmaximum phase, some of the absorption lines corresponding to lower excitation potentials are enhanced and the spectrum is that of a later subtype. This points to a reduction in the degree of excitation in the layer where the absorption lines originate. The degree of ionization therefore also decreases. For example, the equivalent width of the

λ 4227 Å Ca I line in the spectrum of o Cet is more than tripled between the maximum and the minimum, whereas the equivalent widths of the H and K lines of ionized calcium Ca II are reduced to one third during the same period.

Other absorption lines, in particular some Fe I lines, grow weaker along the descending branch of the light curve, and are eventually replaced by bright lines, which constitute the most prominent feature in the spectra of long-period variables between the maximum and the minimum. These bright lines reach peak intensity long before the minimum of the light curve, and subsequently their intensity rapidly diminishes, so that only the strongest remain visible at the minimum, e.g., λλ 4202 and 4308 Å Fe I lines and the λ 4571 Å Mg I line. After the minimum, these lines also disappear.

Approximately half-way between the minimum and the maximum on the rise branch, the spectra of long-period variables again develop bright lines, but now these are emission lines of hydrogen and ionized atomic iron. They reach peak intensity at the maximum, and then gradually grow weaker and disappear near the minimum. High-dispersion spectrograms show that at least some of the long-period variables show bright H and Fe II lines at the minimum also.

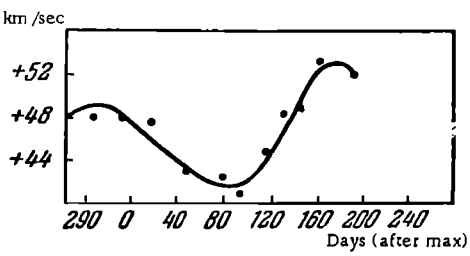


FIGURE 90

Some time after the maximum, the spectra of numerous long-period variables show bright [Fe II] lines. They grow stronger and just before disappearing their intensity is greater than that of the allowed Fe II lines.

Bright hydrogen lines are initially fairly wide (0.5—1.0 Å). Subsequently they grow narrower. Some of these lines appear composed of several components.

The shift of the bright H and Fe II lines is a function of the phase. The corresponding curve for o Cet is shown in Figure 90. The largest violet shift is observed at the maximum. The shift of the bright lines decreases toward the minimum.

The shift of the absorption lines does not follow such a regular curve, and in general the shift amplitudes are smaller than those of the bright lines. The difference in the shifts of the various absorption lines at any given time often exceeds the shift amplitude of individual lines.

The observed Balmer decrement in the spectra of long-period Me variable stars is markedly different from the Balmer decrement of other objects with bright spectral lines. Whereas in the spectra of novae, Be stars, and so on, the H_α line is the strongest in the Balmer series, each successive line showing a lower intensity, the spectra of long-period Me variables show no H_α emission altogether, and it is H_δ that remains the strongest up to the maximum. In the postmaximum stage, H_γ dominates, whereas H_β is invariably weaker than either H_γ or H_δ. Similar anomalies are observed in the intensities of some other Balmer lines.

The Balmer decrement in the spectra of long-period variables of types Se and Ce is on the whole similar to that in Be stars and other stars with bright spectral lines.

The Balmer decrement anomaly in the spectra of Me long-period variables was explained by Shain /6/, who attributed the anomaly to differences in the screening of Balmer line radiation by TiO molecules. The effect of TiO absorption is particularly pronounced in the region of the spectrum where H_α and H_β lines are situated, whereas the spectral region containing the H_δ and higher lines of the Balmer series is comparatively free from TiO absorption. Later Shain established that the anomalies in the intensities of the bright Fe II lines (in particular, the unusual distribution of intensity between the components of the λλ 4233, 4179, 4173 triplet) were due to partial absorption of the radiation at these wavelengths by Fe I and Ca I atoms /7/. These atoms thus occur in atmospheric layers which lie above the layers where the bright lines originate. Joy /8/ also found that the splitting of some bright hydrogen lines into components was due to superposition of metallic absorption lines onto the bright lines.

The H and Fe II bright lines are veiled particularly effectively right after their appearance in the spectrum (before the maximum). As the spectrum develops, the absorption of radiation in these lines by molecular and atomic metals diminishes. Half-way down the descending branch, this effect becomes insignificant. Soon before the minimum, the H and Fe II bright lines are themselves superimposed on molecular absorption bands /9/. The anomalously low intensity of H_β and the total absence of H_α are apparently associated with the absorption of this radiation by atomic hydrogen (in the second state) in the outer atmospheric layers, since according to Shain /7/ the continuum near H_α definitely shows in the spectrum.

The variation of the Balmer gradient in Se and Ce long-period variables has not been studied, and it can be compared with the Balmer gradient in the spectra of Me stars only at epochs near the maximum. At this phase the Balmer decrement in the spectra of Me long-period variables, corrected for molecular absorption bands, is similar to that in the spectra of Se and Ce stars and other objects with bright spectral lines. The bright lines in the spectra of Be stars, Wolf-Rayet stars, and planetary nebulae are the result of degradation of high-frequency stellar radiation into radiation of lower frequencies. This mechanism, however, cannot explain the origin of the bright lines in the spectra of long-period variables, as their temperature is too low (judging from their color and absorption spectrum). Shain appropriately advanced a hypothesis suggesting an excess of short-wave radiation in long-period variables /10/. This hypothesis, however, has not been confirmed so far.

A different explanation of the discrepancy between the low-temperature absorption spectrum and the bright spectral lines was offered by Sobolev, who described the long-period variables as hot giants surrounded by an envelope of high optical thickness in the visible spectrum ($\tau > 1$) /11/. Bright lines originate in the interior of this envelope, where the hydrogen is ionized by the radiation from the star, whereas the outer layers, which do not receive any high-frequency radiation, produce the low-temperature absorption spectrum. The intensity of the bright lines in the spectrum of a hot star surrounded by a dense envelope and the color temperature calculated from the theory of moving stellar envelopes show a good fit with observations near the maximum of the light curve.

The existence of envelopes around long-period variables is supported by observational findings. Various facts suggest that long-period variables eject stellar matter /10/. One of these indicators is the large K effect, measured from the shift of the bright lines at the maximum; the K term is about 15 km/sec. On the other hand, the K term determined from the shift of the absorption lines at the same epoch is nearly zero. Hence it follows that the shift of the absorption lines at the maximum approximately corresponds to the line-of-sight velocity of the star, so that the velocity of the layers where these lines originate relative to the star is small. Deeper lying atmospheric layers, however, where the bright lines originate, move away from the star, i.e., stellar matter is ejected which builds up a moving envelope. The ejected matter is apparently decelerated, and the envelopes show a radial velocity gradient.

The envelopes of long-period variables are nonstationary. This is evident from the radical changes in the spectrum of these stars along the descending branch. As we have noted above, the outer layers of the envelope grow more transparent, but the absolute intensity of the hydrogen emission lines decreases; the main contribution to the low-termperature absorption spectrum now comes from the interior parts of the envelope. Therefore, the parts of the envelope which contained ionized hydrogen in the premaximum phase contain ground-state hydrogen at the minimum. The envelope is thus de-excited: it radiates all the stored energy, without receiving any new energy. The source of ionization apparently ceases to operate near the maximum. *

Let us briefly consider the effects associated with the de-excitation of the envelope and the corresponding interpretation of the spectra of long-period variables in the postmaximum epoch /12/

Before and during de-excitation, the interior parts of the envelope emit high-intensity radiation, in particular in the Lyman lines. Reaching the outer layers of the envelope, this radiation ionizes metals and lowers the absorbance of these layers; the veiling of H and Fe II lines is reduced, while the absorption lines of metals disappear. Recombinations of neutral metal atoms produce bright Fe I, Mg I, and other lines.

When the de-excitation of the envelope in hydrogen lines is completed and a smaller amount of metal ionizing radiation reaches the outer layers, the de-excitation of the latter begins in neutral metal lines. The intensity of these lines decreases, and at the minimum (or soon after) they disappear. The de-excitation time of the atmosphere in the lines of some element (t_n) is given by

$$t_n \approx \frac{1}{Cn_e^0}, \qquad (1)$$

where C is the recombination coefficient, and n_e^0 is the number of free electrons in 1 cm^3 (equal to the total number of metal atoms). The envelopes of long-period variables contain 10^6-10^7 metal atoms in 1 cm^3. Therefore $C \approx 10^{-13}-10^{-14}$, and we find t_n of the order of 100 days, which in fact corresponds the observed lifetime of the bright metallic lines. The recombination coefficients of different metals are different, and the bright lines therefore do not disappear all at once.

* Recently the motion of the layer responsible for the Balmer line emission and the ionization of hydrogen in this layer have been explained /35/ by the action of a shock wave propagating outward through the shell. (Note added in proof.)

As the de-excitation of the ionized hydrogen layer advances, the absolute intensities of the Fe II and [Fe II] lines originating in the same layer decrease. The emission in these lines is mainly excited by collisions with free electrons. The increase in the relative intensity of forbidden lines is associated with the decrease in the free electron density (n_e) in the de-excitation layer. This can be seen from the intensity ratio of the atomic lines. For simplicity, assume that the atom only has three energy levels (the transition between the first, ground level and the second level is forbidden).

If the medium is transparent to line radiation, the ratio of the intensity of the allowed line I_{32} to the intensity of the forbidden line I_{21} is given by

$$\frac{I_{32}}{I_{21}} \approx \frac{\nu_{23}}{\nu_{12}}\left(1+\frac{a_{21}}{A_{21}}\right)e^{-\frac{h\nu_{23}}{kT_e}}, \qquad (2)$$

where a_{21} is the "deactivation coefficient" of excited atoms in state 2, A_{21} is the Einstein coefficient of the 2→1 spontaneous transition, ν_{12} and ν_{23} are the line frequencies, and T_e is the electron temperature. Since $\frac{a_{21}}{A_{21}} \sim n_e$, a decrease in n_e involves a decrease in the ratio $\frac{I_{32}}{I_{21}}$.

The known values of a_{21} and A_{21} for Fe II atoms and the observed relative intensities of the forbidden lines in the spectrum of o Cet show that the n_e in the de-excitation layer varies from 10^{11} near the maximum to 10^7 at the minimum.

The nonstationary effects in the atmospheres of long-period variables should be closely related to the light variation of these stars. The light variation of long-period variables has not been satisfactorily explained to this day. Some authors tried to apply the pulsation theory to these stars (Scott /13/, Zhevakin /14/). However, the principal spectroscopic features of long-period variables do not fit in the framework of the pulsation hypothesis, which at least in its usual form is found to be inconsistent for these stars. The sudden jumps in the magnitude of the long-period variables, mentioned in the preceding, are in outright contradiction with the pulsation theory.

The theory of stellar variation cannot be limited to consideration of atmospheric processes: it should also cover various aspects associated with the interior stellar structure. A satisfactory theory of light variation is therefore considerably more complex than a simple interpretation of effects taking place in the outermost layers of the atmosphere, where the line spectrum is formed. The spectroscopy of long-period variables seems to indicate that ejection of stellar matter plays an important role in light variation and spectrum variation processes of these stars.

TABLE 2

Period in days	n	V, km/sec
150—199	27	129±26
<150 and 200—299	124	42±7
300—349	75	14±8
>349	79	7±3

In conclusion of this section let us consider the motion of long-period variables and their absolute magnitudes. A study of the motions of long-period variables /15, 16/ has revealed a statistical correlation between the group velocity and the period. Table 2 lists the group velocity relative to the Sun (V) for stars with various periods P (n is the number of stars).

The differences in the kinematic properties of various groups of long-period variables points to a definte inhomogeneity of the entire population. Kukarkin /17/ studied the distribution of long-period variables in space and came to the conclusion that the subsystem of these stars on the whole occupies an intermediate position between the halo component (e.g., the system of RR Lyrae stars) and the disk component (e.g., the subsystem of long-period cepheids) of our Galaxy.

The determination of the absolute magnitudes of long-period variables from their proper motions reveals that large light-variation periods correspond to small visual absolute magnitudes ($\overline{M}_v$) at the maximum. Table 3 lists the values of $\overline{M}_v$ for various periods /15/.

TABLE 3

$\overline{P}$, days	150	175	200	250	300	350	400	450	500
M_v	$-2^m.2$	$-2^m.7$	$-2^m.2$	$-1^m.4$	$-0^m.7$	$-0^m.2$	$+0^m.3$	$+0^m.6$	$+0^m.9$

According to other data /18/, the average absolute magnitude of stars with periods over 300 days is substantially higher, being on the average $-1^m.8$. Absolute magnitudes of the same order ($\overline{M}_v = -0^m.5$) are obtained for long-period variables with known trigonometric parallaxes /15/. Hence, these stars should definitely be regarded as giants or supergiants.

The determination of the bolometric luminosity of long-period variables from their visual absolute magnitudes involves considerable difficulties at present, since the application of usual bolometric corrections is highly doubtful in this case. Nevertheless, it seems that the bolometric luminosity of long-period variables is exceedingly high and probably corresponds to the luminosities of early-type supergiants (O and B).

§48. **Semiregular variables — RV Tauri stars.** A distinctive feature of RV Tauri stars is the alternation of deep and shallow minima on the light curve (Figure 91 shows the light curves of three stars of this type: V Vul (a), U Mon (b), and R Sct (c)). Sometimes these minima interchange their positions. This light curve is preserved most of the time, and the deviations from the usual shape are not durable. If the star shows an alternation of deep and shallow minima, but the irregularities in its light curve are longer and more prominent, the star is not classified as a RV Tauri star: it is generally regarded as a semiregular variable of some other type, in particular a yellow semiregular variable (see §49).

Since the classification of variable stars is often based on a comparatively short section of the light curve, many stars without the characteristic light curves and spectra of RV Tau stars have been erroneously classified as such. Only a few dozen stars are definitely known to be RV Tau stars. The brightest of these stars is R Sct ($6^m.1 - 8^m.6$), and RV Tau itself is the fourth brightest star ($9^m.8 - 13^m.3$). The light variation amplitudes of other stars of this type are of the same order or somewhat less ($0^m.8 \leqslant A \leqslant 3^m.5$).

The time interval between two successive minima of the light curve is called the fundamental period, and the time between two deep (primary) minima is the formal period. The formal periods of

RV Tau stars fall between 30 and 150 days. Some stars of this type show cyclic fluctuations of the average magnitude. The length of the cycle is different for different stars, mostly reaching a few hundred days; there are also stars with cycles of up to two thousand days.

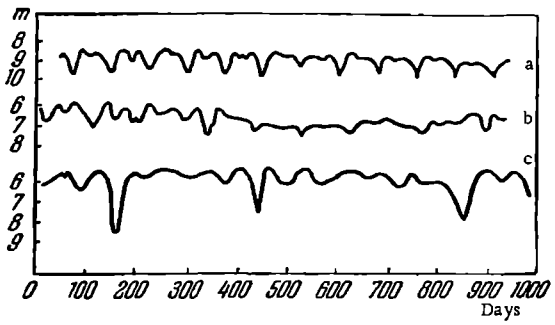


FIGURE 91

The average spectrum of RV Tau stars corresponds to spectral types F, G, or K /19/. The spectra are variable. The intensities of the spectral lines, their shifts and profiles change in phase with the light variation of the star. This indicates that the color temperature of the star is a function of its phase.

The earliest spectrum of RV Tau stars is observed in the epoch half way along the rise branch of the light curve (after the primary minimum). At that time the absorption lines of neutral metals are weak, and the lines of ionized metals are enhanced. On the whole, the spectrum of an RV Tau spectrum at that time greatly differs from the spectrum of a normal star of the same type; some lines in the spectra of RV Tau stars have anomalous intensity. As the magnitude increases, the degree of excitation in the atmosphere diminishes and the spectrum approaches a "normal" spectrum.

Right after the primary minimum, the spectra of numerous RV Tau stars show bright hydrogen lines, which reach peak intensity before the maximum. The hydrogen absorption lines in these epochs are less prominent.

RV Tau stars markedly differ from one another in the intensities of the bright hydrogen lines. For instance, the spectrum of R Sct almost always shows bright lines, whereas in most stars, AC Her in particular, they are observed only on the rise branch. Some stars show no bright hydrogen lines altogether.

Before the primary minimum, the spectra of RV Tau stars often develop TiO absorption bands. The spectral type of the star from absorption lines at this epoch is G5—K0. Stars without bright hydrogen lines in their spectra show no TiO absorption bands either. The spectra of RV Tau stars also show absorption bands of other molecules. For example, the spectrum of AC Her contains strong CN and CH bands; the CH bands appear a few days after the maximum, when the star is of spectral type F2—F5 (judging from the absorption lines).

We see from the above that the early-type features in the spectra of RV Tau stars (bright lines) combined with characteristic features of M and N type spectra (TiO, CN, and other absorption bands). A similar situation was observed for the long-period variables. Neither the bright lines nor the strong molecular absorption bands are typical features of the normal F2—K3 subtypes, to which the metal absorption spectrum corresponds.

The complex spectra of RV Tau stars apparently can be attributed only to the unusual atmospheric structure. At this point we should consider the highly important data on the atmospheres of RV Tau stars obtained by Abt /20/. Until recently, the spectra of RV Tau stars were observed with insufficient dispersion and for this reason the absorption line shifts were presumably seen to vary periodically with the periodic light variation of the star. Since the "radial velocity curve" obtained from these line shifts was similar in its shape to a displaced light curve of the star, both the line shifts and the light variation of RV Tau stars were generally explained using the pulsation theory. However, high-dispersion spectra of U Mon and two other stars led Abt to the conclusion that the continuous "radial velocity curve" is purely imaginary. Actually, the absorption line shifts vary as shown in Figure 92 (the figure gives the light curve and the corresponding line shifts for U Mon only). Most of the time the absorption line is made up of two components with different shifts. Each of these components forms soon after the minimum and disappears after the next minimum. The newly formed component occupies the short-wave position, but since its shift steadily decreases, it develops into a long-wave component before its disappearance.

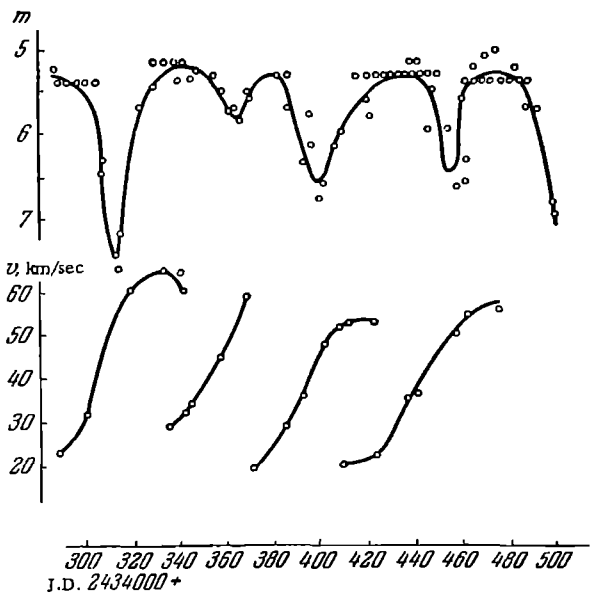


FIGURE 92

Thus, for most of the time U Mon and other RV Tau stars have two absorbing layer. Both layers are decelerated, but they move with different velocities. This fact may be explained by periodic shedding of a gaseous envelope, but the exact mechanism of this phenomenon, and the related mechanism of light and spectrum variation in RV Tau stars, is still unknown.

The "radial velocity" curves of RV Tau stars are similar to those of long-period cepheids and W Vir stars. On the other hand, the presence of bright hydrogen lines near the maximum and TiO absorption bands at the minima links these spectra with the spectra of long-period variables. RV Tau stars thus occupy an intermediate position between W Vir stars and long-period variables. Their periods and (as we shall see below) their luminosities also occupy an intermediate place.

We have already observed that RV Tau stars are nonuniform in their spectral characteristics. A similar nonuniformity is observed in the motions of these stars /21/. RV Tau stars (and the related yellow semiregular variables) fall into two distinct groups in terms of the spatial velocity: the first group comprises slow moving stars (with spatial velocities less than 70 km/sec), and the other group includes fast moving stars with velocities greater than 70 km/sec (sometimes exceeding 200—250 km/sec). The slow stars differ from the fast moving stars in their spectra, specifically in the higher intensity of ionized metal absorption lines and CH absorption bands. The spatial distribution is also different for stars of these groups. Slow stars show a higher galactic concentration (mean galactic latitude 13°) than the fast moving stars, whose mean latitude is 27°. The average absolute photographic magnitude of RV Tau stars determined from the intensity of the interstellar absorption lines and by other methods /21/ is $-3^m.0$ for slow stars and $-1^m.5$ for fast moving stars. These magnitudes clearly fall between the average absolute magnitudes of the long-period cepheids and those of the long-period variable stars.

The above facts suggest that RV Tau stars include stars of different stellar populations.

§49. **Semiregular and irregular variable stars of other types.** RV Tau stars, as we have noted, sometimes show irregular variation of brightness. There are stars whose light curves are on the whole similar to those of RV Tau stars, but the irregularities occur more often and the corresponding periods may be substantially longer. These are stars of spectral types F, G, or K, and are therefore called yellow semiregular variables. They are akin to RV Tau stars in a number of spectral features; in particular, the spectra of these stars often show bright hydrogen lines near the maximum.

For lack of exact criteria differentiating between RV Tau stars and yellow semiregular variables, the classification is somewhat arbitrary.

The alternation of deep and shallow minima is also observed for AF Cyg (Figure 93), a variable star of spectral type M, and for some other variables of this type. These stars are classified as red semiregular variables; they further include a number of other varieties of variable stars, such as RS Cnc stars, μ Cep stars, Z Aqr stars, and others. All the red semiregular variables are of spectral type M; their brightness varies periodically with an amplitude not exceeding 2^m and the length of period ranges from a few tens to a few hundreds of days. The light curves of red semiregular variables show different degrees of irregularity. Some

of these are almost as regular as the long-period variables (Z Aqr, X Mon), whereas others have highly irregular light curves.

FIGURE 93

The light curve of some red semiregular variables appears to be a superposition of two oscillations. For example, the light curve of μ Cep is regarded as an oscillation with a period of a few hundred days of amplitude 0".5 on which oscillations with amplitude 0".1 and period of some 100 days are superimposed. A similar "dual periodicity" is noted for α Ori and other stars. However, the application of statistical methods to a 54-year series of observations of μ Cep /22/ has shown that the light variation of this star is probably an outcome of a stochastic, and not a harmonic, process. The conclusions concerning "dual periodicity" and regular light variation of red semiregular variables should therefore be considered with some reservation.

The spectra of red semiregular variables in the main are no different from the spectra of normal M-type giants and supergiants /23, 24/. Some of these stars (e.g., μ Cep /24/) may show low-intensity hydrogen emission lines on the rise curve. The TiO absorption bands, observed in the spectra of all M-type stars, increase in intensity with decreasing brightness of μ Cep. During one cycle, the absorption line shifts change by a few km/sec.

According to Grigoryan /25/, the radiation of μ Cep is polarized. The degree of polarization varies between 1.0 and 2.3%. No correlation was observed between polarization and light variation (in blue light). The polarization of the light from μ Cep may be indicatory of a nonthermal component in the radiation of this star.

The light curve of semiregular variables may be formally fitted with a superposition of two or more cyclic oscillations. Some stars, however, show completely irregular light curves. These are the so-called irregular variables. Figure 94 is the light curve of one such irregular variable, TU Aur.

FIGURE 94

Almost all irregular variables are stars of spectral type M. They apparently constitute a continuation of the sequence of red semiregular variables in the direction of increasing irregularity of light variation.

All the red semiregular and irregular variables are high-luminosity stars /26, 27/. Some of them (about 10%) are supergiants with absolute magnitudes between $-4^m.5$ and $-2^m.0$. The remaining stars are giants with mean absolute magnitude of about $-0^m.9$.

§ 50. **Spectroscopic variables.**

a) Be stars. Numerous B-type stars with bright spectral lines (the so-called Be stars) show irregular variations of magnitude with an amplitude of $0^m.1-0^m.2$. Larger light variations were noted in isolated cases. For example, the brightness of γ Cas in 1936 increased by $1^m.5$ and after that it fluctuated between $0.5-1^m.0$, while decreasing on the whole (Figure 95).

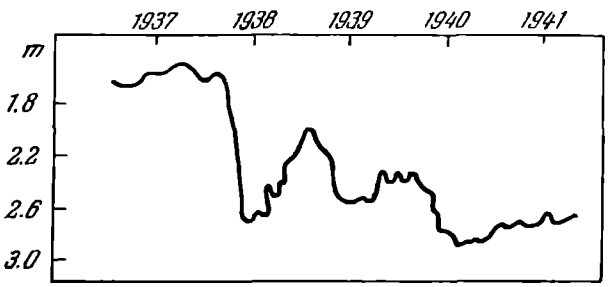


FIGURE 95

The luminosities of Be stars are no different from those of normal B stars, which, as we known, have exceedingly high luminosity. Thus, in accordance with the usual classification, Be stars are high-luminosity irregular variables. The processes in the outer layers of these stars were more often observed and theoretically studied in greater detail than the atmospheric phenomena in other variable stars.

The main feature in the spectra of Be stars, which distinguishes them from the spectra of B stars, is the presence of bright Balmer lines of hydrogen and weaker lines of neutral and ionized metals, O II and C II ions, etc. The spectra of early Be stars (Be0—Be1) also show He I and Fe III lines. The emission line spectrum of Be stars indicates that the bright lines are produced by recombinations of ionized atoms to the higher energy levels followed by cascade transitions to the lower levels (a situation not unlike that leading to the formation of bright lines in the spectra of planetary nebulae). Such processes are possible only in a low-density medium. Be stars thus must have an ionized envelope. This conclusion is supported by observations of the continuous spectrum of Be stars.

Barbier and Chalonge with co-workers carried out a detailed comparison of the energy distribution in the continuous spectrum of Be stars and normal B stars of the same ionization temperature. They showed that on the average the spectrophotometric temperature of Be stars is somewhat less than the temperature of stars without bright spectral lines /28/. Later Tsoi Dyai O /29/ found that this difference between B and Be stars persists even if interstellar absorption is corrected for. There are further differences in the Balmer discontinuity in the spectra of Be and B stars.

In B stars the intensity I on the short-wave side of the Balmer limit (in the Balmer continuum) is less than on the long-wave side (in the Paschen continuum). The Balmer discontinuity D, defined by

$$D = \lg \frac{I_{\nu<\nu_2}}{I_{\nu>\nu_2}}, \tag{3}$$

where ν_2 is the frequency of the Balmer limit, is positive for these stars. For Be stars, D is on the average less than for B stars of the same subtype /28, 29/. The Balmer discontinuity sometimes becomes negative, i.e., near the Balmer limit the radiation in the Balmer continuum is stronger than in the Paschen continuum.

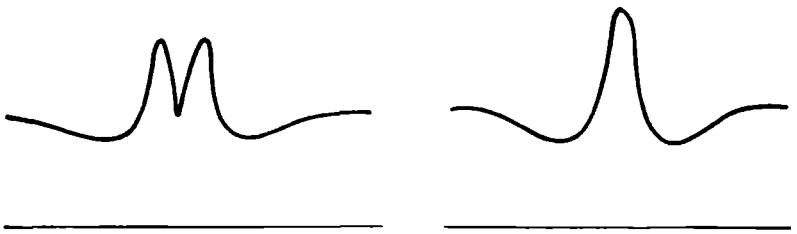
The peculiar features of the Be spectrum can be interpreted as a superposition of the spectrum of the envelope on the spectrum of a B-type star. The radiation of the envelope in the Balmer continuum is stronger than in the Paschen continuum and the D of a star with an envelope is therefore less than that of a star without an envelope. Moreover, the energy distribution in the spectrum of the envelope corresponds to a much lower temperature than the electron temperature of the envelope, which in its turn generally does not exceed the temperature of the star. The superimposed radiation of the envelope therefore generally lowers the color temperature of the star.

This is the qualitative interpretation of the continuous spectrum of Be stars. More detailed discussion of these problems and other aspects of the theory of Be stars will be found in the chapter on stellar atmospheres (Chapter XV). Note that comparison of theory with observation for a number of Be stars made it possible to determine the contribution of the envelope to the total radiation of the star, the electron temperature of the envelope T_e, and the annual mass ejection of the star /30, 31/. Beyond the Balmer series, the radiation of the envelopes accounts for about 0.1 of the total radiation of the star. T_e is equal to 30,000° for B0e stars and 10,000—20,000° for B1e—B3e stars. A Be star ejects yearly a mass of the order of 10^{-6}—10^{-7} solar masses. In 1936—1938, γ Cas showed a stronger ejection of matter, and at that time its envelope was larger than that of other Be stars.

Envelopes of Be stars are opaque to radiation in hydrogen Balmer lines, and the observed intensities and profiles of bright lines are therefore explained using the theory of moving envelopes /11, 30/.

Let us consider the line profiles in the spectra of Be stars. In most cases the bright lines appear superimposed on wide and shallow absorption lines and are split into two components (Figure 96a). Stars with one-component bright lines are also sometimes observed (Figure 96b). The width of the bright lines is a few angstrom, whereas the underlying absorption lines are substantially wider (sometimes reaching 20 Å).

The bright lines in the spectra of Be stars are subject to variations. Mainly the line profiles change, and the equivalent widths are conserved. The relative intensities of the bright line components are generally variable. The "central depression" is shifted toward the stronger component. In a number of cases, the equivalent width of the bright lines was also known to change. Occasionally it gradually decreases until the bright lines disappear altogether. After some time, the lines reappear. No regularity was observed in these disappearances and reappearances of bright lines.



a b

FIGURE 96

To interpret the line profiles in the spectra of Be stars, we should remember that at times when the spectrum shows no bright lines, the stars superficially do not differ from normal B-type stars. The wide and shallow absorption line profiles in numerous B stars are attributed to the fast spin of these stars. Be stars apparently also spin, entraining their envelopes at a high velocity. The spin velocity of the envelope at the effective level of line emission is of course less than at the surface of the star (because of angular momentum conservation in the envelope), and the width of the bright lines is therefore less than the width of absorption lines.

The envelopes of Be stars are the result of ejection of stellar matter. This ejection amounts to steady flow (since the emission features of the spectrum are fairly stable) and should produce an extended envelope. These considerations are confirmed by the likeness in the theoretical profiles of bright lines originating in a spinning and expanding envelope and the line profiles observed in the Be spectra /30/.

The recurrent disappearance of the bright lines from the spectra of numerous Be stars can only be attributed to a temporary stop in ejection, which is thus not quite steady. The nonstationary character of ejection also follows from the variation of the bright line profiles /30/, which do not involve any change in the equivalent width.

When the ejection diminishes, without stopping altogether, the surrounding envelope becomes less dense and its transparency increases, whereas if ejection is intensified the density of matter in the envelope and the optical thickness both increase. If the envelope remains opaque to line radiation, the variation of its density hardly affects the energy emitted in the relevant lines. To check this point, consider a Balmer line of frequency ν_{2k}. The energy E_{2k} radiated by the extended envelope in this line is given by the theory of moving stellar envelopes:

$$E_{2k} = 4\pi A_{k2} h\nu_{2k} \int_{r_0}^{\infty} n_k \beta_{2k} r^2 dr, \qquad (4)$$

where n_k is the number of atoms in a state k in 1 cm^3, β_{2k} is the fraction of quanta emerging from the envelope due to the velocity gradient, and r_0 is the radius of the star. The other symbols in (4) are conventional.

Since $\beta_{2k} \sim \frac{1}{n_2}$, we have $n_k \beta_{2k} \sim \frac{n_k}{n_2}$. The degree of excitation, and hence $\frac{n_k}{n_2}$, is not particularly sensitive to the density of the envelope, and therefore

changes in ejection intensity leave E_{2h} approximately constant. At the same time, the distribution of intensity inside the line (the line profile), which is determined mainly by $n_h(r)$ and the velocity of the envelope (see Chapter XVII), essentially depends on the rate of ejection of stellar matter. Thus, variations in the bright line profiles without changes in the equivalent line width are indicatory of changes in the rate of ejection of stellar matter.

When the envelope density changes (due to the nonsteady-state ejection from the star), the continuum radiation of the envelope is altered. The observer always receives the total radiation of the star plus its envelope. When the radiation of the envelope changes, the observed brightness of the star will also change.

Let Δm be the difference in magnitudes of the star without an envelope and a star surrounded by an envelope. The optical thickness of the envelope in the visible spectrum is generally much less than unity (after all, absorption lines which belong to the reversing layer of the star proper are seen without considerable distortion through the envelope). Δm is therefore given by the simple expression

$$\Delta m = 2.5 \lg \frac{E_* + E_{\text{env}}}{E_*}, \tag{5}$$

where E_* is the energy radiated by the star without the enveloe, E_{env} is the energy radiated by the envelope at some effective wavelength. (For example, if Δm is the difference in visual magnitudes, we may take λ_e = = 5550 Å).

Since the envelope emits following recombination of hydrogen atoms, we can use the standard expression for the energy radiated by unit volume (see Chapter XVII) and integrating over the entire volume of the envelope we get

$$E_{\text{env}} = R(T_e)\, e^{-\frac{h\nu_0}{kT_e}} \int n_e n^+ \, dV, \tag{6}$$

where $R(T_e)$ is some known function of electron temperature, n_e and n^+ is the number of electrons and protons in 1 cm^3, respectively. When the rate of ejection changes, the density of matter in the envelope, and hence the product $n_e n^+$, are affected. E_{env} therefore changes. If the energy E_* radiated by the star remains constant, Δm and the apparent brightness of the star should change correspondingly. In most cases we can hardly expect the magnitude of the star itself to vary with time, and the light variation is therefore attributable to changes in the luminosity of the envelope, as in the case of γ Cas /30/. In certain cases, however, the brightness of the star itself changes (the decrease in the brightness of Pleione in 1936).

The mechanism of light variation of Be stars has thus been established to first approximation. However, the nature of Be stars, and in particular the reasons for the irregular ejection of matter are not clear at this stage.

b) Magnetic variables — α^2 CVn stars. α^2 CVn stars belong to spectral type A and are characterized by strong lines of rare earth elements in their spectra. Their luminosities are higher than those of main-sequence stars of the same type by about 1^m.

Belopol'skii found that the intensities of various lines in the spectrum of α^2 CVn underwent periodic changes. These lines fell into two groups (the

Eu II group and the Cr II group), with the intensity varying in opposite directions. The line shifts and line profiles varied together with the line intensities.

Photoelectric observations established light variation of α^2 CVn stars. Thus, in particular, the brightness of α^2 CVn varies by $0^m.05$ in yellow light and by $0^m.03$ in blue light /32/. The light variation period is equal to the period of line intensity variation.

Babcock established that stars of this type possessed strong magnetic fields /33/. The field of α^2 CVn periodically reverses with an amplitude of 4000 gauss.

The variation of line intensities and line shifts in the spectrum of HD 125248 (another α^2 CVn star) is compared in Figure 97 with the variation of its magnitude and its magnetic field at the poles.

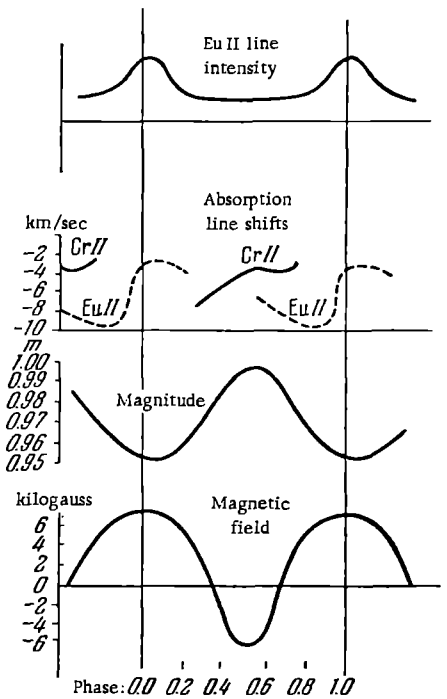


FIGURE 97

The relationship between spectrum and light variation, on the one hand, and the magnetic field variation on the other is obvious. Deutsch established a correlation between the widths of the unchanging lines and the period of the star /34/. He thus came to the conclusion that α^2 CVn stars were spinning stars observed in or near the equatorial plane. The period of spectrum variation is in fact the period of rotation. The magnetic and spectrum variations are explained by the assumption that the magnetic axis of the star is inclined to the spin axis. The light variation, if this assumption is true, can be explained by nonuniform brightness of the different surface zones.

REFERENCES TO CHAPTER VIII

1. LUDENDORFF, H.— Stzb. Preuss. Akad. Wiss., XX. 1932.
2. HETZLER, Ch.— ApJ, 83:372. 1936.
3. PETTIT, E. and S. NICHOLSON.— ApJ, 78:320. 1933.
4. MAYALL, M.— J. Roy. Astron. Soc. Canada, 50:135. 1956.
5. MERRILL, P. Spectra of Long-Period Variable Stars.— Chicago. 1940.
6. SHAIN, G.A.— Zs. f. Ap., 10:73. 1935.
7. SHAIN, G.A.— Doklady AN SSSR, 42:254. 1944; 44:293. 1944.
8. JOY, A.— ApJ, 106:288. 1947.
9. JOY, A.— ApJ Suppl. Series, 1:39. 1954.
10. SHAIN, G.A.— Izvestiya AN SSSR, phys. ser., 9:161. 1945.
11. SOBOLEV, V.V. Dvizhushchiesya obolochki zvezd (Moving Stellar Envelopes).— Leningrad. 1947.
12. GORBATSKII, V.G.— Astron. Zh., 34:860. 1957; 35:748. 1958.
13. SCOTT, R.— ApJ, 101:71. 1945.
14. ZHEVAKIN, S.A.— Astron. Zh., 31:335. 1954.
15. WILSON, R. and P. MERRILL.— ApJ, 95:248. 1942.
16. MERRILL, P.— ApJ, 94:171. 1941.
17. KUKARKIN, V.B. Issledovanie stroeniya i razvitiya zvezdnykh sistem na osnove izucheniya peremennykh zvezd (Structure and Evolution of Star Systems from a Study of Variable Stars).— Gostekhizdat. 1949.
18. OORT, J. and J. Van TULDER.— BAN, 9:332. 1939.
19. ROSINO, L.— ApJ, 113:60. 1951.
20. ABT, H.— Astron. J., 58:210. 1953; ApJ, 122:72. 1955.
21. JOY, A.— ApJ, 115:25. 1952.
22. ASHBROOK, J., R. DUNCOMBE, and A. Van WOERKOM.— Astron. J., 59:12. 1954.
23. KEENAN, Ph.— ApJ, 95:461. 1942.
24. McLAUGHLIN, O.— ApJ, 103:35. 1946.
25. GRIGORYAN, K.A.— Bull. Byurakan Observ., 25:45. 1958.
26. JOY, A.— ApJ, 96:344. 1942.
27. WILSON, R.— ApJ, 96:371. 1942.
28. BARBIER, P. and P. CHALONGE.— Ann. d'ap., 4:1. 1941.
29. TSOI DYAI, O.— Astron. Zh., 33:506. 1956.
30. GORBATSKII, V.G.— Astron Zh., 26:307. 1949; 28:450. 1951.
31. TSOI DYAI, O.— Astron. Zh., 33:628. 1956.
32. PROVIN, S.— ApJ, 118:489. 1953.
33. BABCOCK, H.— ApJ, 110:126. 1949.
34. DEUTSCH, A.— Trans. IAU, 8:801. 1954.
35. GORBATSKII, V.G.— Astron. Zh., 38:256. 1961.

V. G. GORBATSKII

Chapter IX

LOW- AND MEDIUM-LUMINOSITY VARIABLES

In this chapter we consider variable stars whose luminosity is low at all phases of the light curve or at least in the minimum (the novae and recurrent novae are dealt with in a separate chapter). Observations of these stars are a hgihly complex undertaking because of the low luminosity and the generally irregular light variation; without sufficiently sophisticated technical equipment, these stars cannot be observed altogether. Our information on the spectra of these stars (and often on their light variation as well) is therefore highly fragmentary and the number of known stars of this type is not large. Nevertheless, the scanty observational findings have led to a number of important conclusions concerning the nature of some of the low-luminosity variables; thus, Ambartsumyan has shown /1/ that T Tau stars cannot be older than a few million years; he has also established the partly nonthermal origin of the radiation of T Tau, UV Cet, and some other variables. Studies of variable dwarfs have thus acquired primary importance for astrophysicists, mainly in connection with the problem of stellar evolution.

§ 51. **Z Andromedae stars.** Z And stars are also known as nova-like variables. This name reflects the considerable likeness between Z And stars and recurrent novae in terms of light variation and some spectral features.

The apparent magnitude of all known nova-like stars is low. One of the brightest is the prototype of the entire group, Z And itself; at the maximum it reaches 8 mag. Z And has been studied to a greater extent than other nova-like stars (AX Per, BF Cyg, AG Peg, etc.) and we will therefore consider the corresponding results in some detail.

Z And is occasionally seen to increase in brightness by a few magnitudes (this is referred to as the "flare"). The flare is followed by fluctuations of brightness with decreasing amplitude. In this respect, the light curves of nova-like variables (the light curve of Z And is shown in Figure 98) are reminiscent of the light curves of some novae, which show damped oscillations on the descending branch (e.g., Nova Persei 1901). The difference between the light curves of novae and nova-like stars is that the rise branch and the subsequent fluctuations in Z And stars are much slower than in novae, and the flare amplitudes are substantially smaller than those of novae.

Comparison of the visual and photographic light curves of Z And shows that the color index CI = $m_{pg} - m_{pv}$ is high at the minimum, reaching about 1^m, whereas with increasing brightness it falls to $0.3 - 0^m.4$. Hence the

conclusion that the spectrophotometric temperature of the star increases with increasing brightness. Measurements of the spectrophotometric temperature T_s of Z And give 5200° ± 900° at the minimum /3/. The T_s of other nova-like stars at the minimum is even smaller. Thus, for CI Cyg $T_s \approx 3700°$ /4/. The Balmer jump in the specra of AX Per and BF Cyg is negative, i.e., the intensity in the Balmer continuum is higher than the intensity in the Paschen continuum. For AX Per $D = -0.4$ and for BF Cyg $D = -0.1$ /5/.

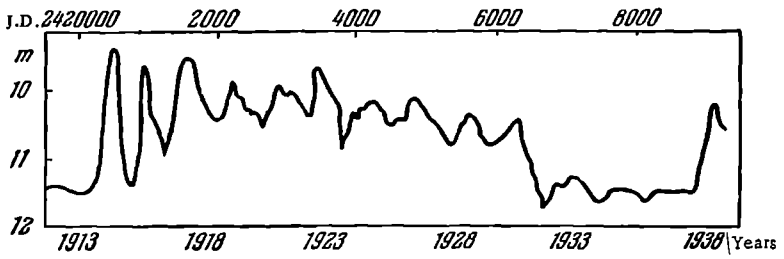


FIGURE 98

The line spectrum of Z And at the minimum /5, 6, 7/ consists of two groups of bright lines: a) lines with high excitation potential, e.g., He II, N III, and other elements, including the forbidden lines [O III], [Ne III], [N V]; these lines are assumed to originate in a nebula (not unlike a planetary nebula), and are correspondingly regarded as the so-called "nebular part" of the emission spectrum; b) lines of lower excitation potential (H, He I, Mg II, Fe II, and others); they apparently form in the stellar atmosphere, and are thus referred to as the "stellar part" of the emission spectrum.

The spectrum radically changes during the flare. The high-excitation lines disappear, and the "stellar" lines are converted to absorption lines with a violet emission components. The spectrum of Z And thus approaches the spectrum of a P Cyg star. This likeness is further accentuated by the shortward shift of the absorption lines (on the average by −80 km/sec), as in the spectra of P Cyg stars. At the maximum, the energy distribution in the spectrum of Z And corresponds to spectral types A—F.

The conversion of the emission spectrum into a P Cyg spectrum during the flare of 1939 took several months. The brightness of the star increased by 3^m, and the spectrophotometric temperature rose from 5000 to 10,000—12,000°. The decline of brightness (after a very gentle maximum) was accompanied by a weakening of the P Cyg absorption lines. The "nebular" emission lines reappeared and gradually grew in intensity. A year after the beginning of the flare, the P Cyg lines vanished; very strong emission in the Balmer continuum was observed, and the spectrum recovered its usual appearance.

During the brightness fluctuations that followed the flare, no P Cyg lines appeared and the high-excitation lines did not disappear entirely; they only grew weaker relative to the "stellar" lines as the brightness increased in each fluctuation.

The flare of Z And was explained by O. Struve and Swings /6/; in their opinion, the flaring star ejected a substantial quantity of stellar matter which formed an extended shell around it. This shell was responsible for the P Cyg spectrum. At the same time, the high-frequency radiation of the star (beyond the Lyman limit), which before the flare excited the outer nebula, was absorbed in the newly formed shell; the high-excitation emission lines therefore disappeared from the spectrum.

As the ejection gradually stopped, the brightness started declining, and the optical thickness of the expanding envelope began decreasing. The P Cyg lines thus grew weaker, and the high-frequency radiation of the star again started exciting the nebula.

The effects attendant on the flare of Z And are thus highly similar to the outburst of a nova; the mechanism of brightness increase is apparently the same in both cases (see Chapter XI), and the brightness fluctuations in the postmaximum stage are associated with occasionally renewed ejection, though with smaller intensity.

Struve and Swings's interpretation of the flare of nova-like stars disregards some important spectral features, primarily the low spectrophotometric temperature of the star (at the minimum); another interesting spectral feature is the presence of TiO absorption bands alongside with the bright emission lines at the minimum. These absorption bands are generally typical of cold M stars. As the brightness increases, these stars grow weaker and eventually disappear; on the descending branch they reappear again.

The spectra of nova-like variables at the minimum are thus a combination of early-type features with some late-type features. To explain this fact, it has been suggested (e.g., /8/) that nova-like variables are binary systems comprising a cold M-type star and a hot star. In Berman's opinion, the bright lines originate in the extended atmosphere of the hot star, and the TiO absorption bands form in the atmosphere of the cold star.

The binary hypothesis of nova-like variables is highly popular, although it leads to a number of obvious objections. Thus, for instance, to ensure a good fit between the hypothesis and the observations, the two system components have to be regarded as variables, with synchronous light variation. This coincidence is odd in the greatest degree and requires a special explanation. Moreover, no direct confirmation of the binary nature of any of the nova-like variables has been obtained so far. Therefore, although it is possible that some nova-like stars are indeed binaries, this phenomenon can hardly be applied to explain the generally complex form of their spectra.

Sobolev developed a radically different approach to the spectra of Z And stars /9/. He suggests that these stars (like the long-period variables) are hot stars surrounded by an envelope. The optical thickness of the envelope in the visible spectrum is fairly high ($\tau > 1$). The interior of the envelope is ionized by the stellar radiation, and it is here that the bright lines form. The high-frequency radiation does not reach the outer layers of the envelope, and they therefore have a comparatively low temperature producing a low-temperature spectrum.

This approach to the spectra of nova-like variables is consistent with the ejection of stellar matter by these stars and on the whole offers a good explanation of their peculiar features. Unfortunately, the presence

of the bright "nebular part" of the spectrum, together with absorption bands, is not accounted for by this theory. As we have mentioned above, the "nebular" spectrum is emitted by a nebula which is relatively distant from the star. Moreover, the observations show that the intensity of these bright nebular lines and the intensity of absorption bands increase simultaneously /10/. Yet if the emission line spectrum and the absorption spectrum are formed in the same envelope, an opposite effect should be expected. The theory therefore should be further developed to account for the variety of these factors.

§ 52. U Geminorum and Z Camelopardalis stars. The brightness of U Gem stars occasionally jumps by $2-4^m$ and then rapidly decreases to the original value. Between these "flares" the magnitude of the star remains fairly constant. The time from the beginning of one flare to the next (the cycle) varies between wide limits for any given star; for instance, SS Cyg (the brightest of U Gem stars) shows cycles ranging from 20 to 110 days. The successive flares of the same star may also have different amplitudes. Figure 99 shows a specimen light curve of U Gem.

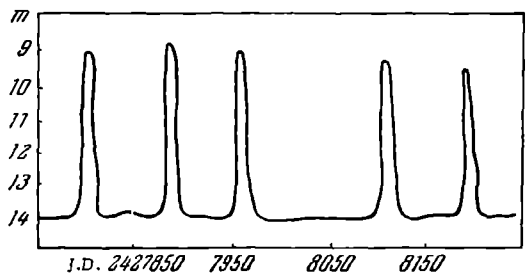


FIGURE 99

On the other hand, each star is characterized by a certain mean cycle and a certain mean flare amplitude.

In 1933 Parenago and Kukarkin discovered a statistical relation between these mean quantities /11/. The longer the mean cycle $\overline{P}$, the greater is the mean amplitude $\overline{A}$. Figure 100 is a plot of this dependence for U Gem stars ($\overline{A}$ is expressed in stellar magnitudes, $\overline{P}$ in days).

A similar dependence between the length of the preceding cycle and the flare amplitude is observed for every individual U Gem star.

The analytical relation between $\overline{A}$ and $\overline{P}$ is expressed by the equality

$$\overline{A} = 0^m.4 + 1.85 \lg \overline{P}.$$

The physical meaning of this dependence is not clear at present. Note, however, that it is observed not only for U Gem stars. A similar relation exists between the cycles and the amplitudes of the recurrent novae. This fact is probably indicatory of a fundamental similarity in the flare mechanism of U Gem stars and recurrent novae. The observed flare, however, is entirely different in U Gem stars and in novae: there is no likeness whatsoever in the light curves of the two groups of stars.

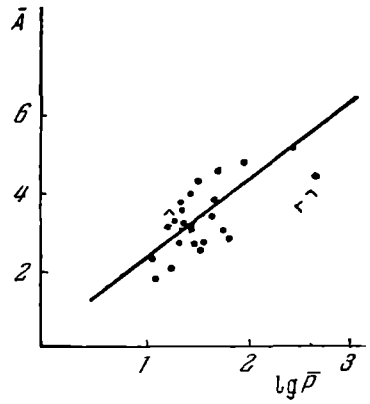


FIGURE 100

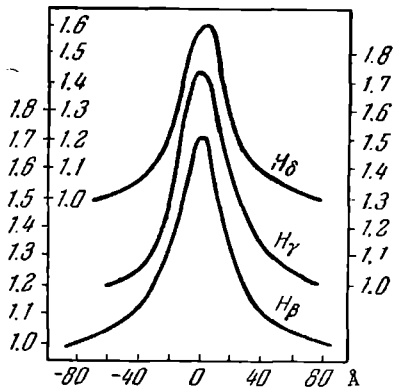


FIGURE 101

Besides the large flares, SS Cyg at and near the minimum shows frequent small fluctuations with amplitude of $0.1-0^m.2$ taking 2—5 min and brief flares with amplitudes of $0.3-0^m.6$ /12, 13/. The rate of light variation reaches $0^m.25$ per min. The absolute amplitude of these fluctuations does not change when the average brightness of the star is increased.

The spectra of U Gem stars /14, 15/ in the postmaximum stage and between successive flares are somewhat reminiscent of the spectra of novae: they show very strong bright Balmer lines and somewhat weaker ionized helium lines. The profiles of the bright lines (Figure 101) are highly peculiar, however, and have nothing in common with the emission line profiles in the spectra of novae. The width of the bright hydrogen lines observed in the spectra of U Gem stars reaches 50 Å; they are symmetric and the shift of their centers from the normal position is not large (1 Å). Together with the bright lines, the spectra of U Gem stars show exceedingly strong emission in the Balmer continuum. In particular, in SS Cyg the intensity of the Balmer continuum near the Balmer limit is a factor of 6—7 greater than the intensity on the other side of the limit.

Between successive flares, the energy distribution in the visible continuum of SS Cyg corresponds to temperatures of 5000—5500°. The color temperatures of other stars of this type are of the same order of magnitude. There are some observations of weak absorption lines corresponding to dG5 spectra at the light minimum of SS Cyg.

As the brightness increases, the color temperature rises reaching 14,000—15,000° at the maximum (which corresponds to a star of type A0). The line and continuum emission disappear at this point. The bright hydrogen lines are replaced by absorption lines of similar width. The absorption line profiles are symmetric and on the whole similar to the line profiles in the spectra of white dwarfs.

In the postmaximum stage, a symmetric bright line develops at the center of each absorption line; its intensity increases as the star declines

to the minimum. Figure 102 shows the line profiles in the spectrum of SS Cyg on the descending branch of the light curve.

The likeness of U Gem stars at the maximum to white dwarfs is not restricted to the absorption spectrum and the energy distribution in the continuum. The luminosities of these stars are also close to the luminosities of white dwarfs. The visual absolute magnitude of U Gem stars at the maximum (determined from their motions) is $+6^m$. Trigonometric parallax measurements of SS Cyg lead to the same result. Hence, U Gem stars at the maximum are typical white dwarfs. The wide absorption lines in the spectra of white dwarfs are generally attributed to the Stark effect. The broadening of the absorption lines in the spectra of U Gem stars is apparently due to the same factor. Anyhow, the considerable width and the peculiar profiles of the absorption (and emission) lines in the spectra of these stars cannot be explained by atmospheric motions.

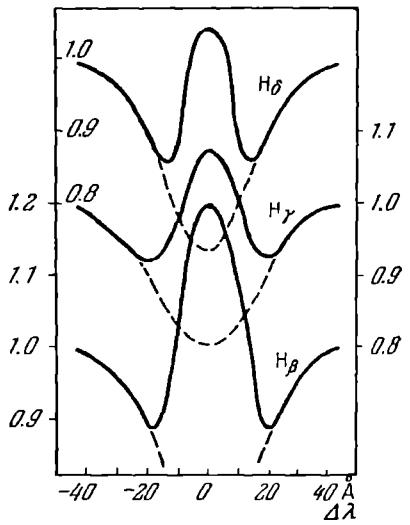


FIGURE 102

Of considerable interest is the origin of the bright lines in the spectra of U Gem stars. Since the spectra of these stars also show Balmer continuum emission and the normal Balmer decrement of a star with bright lines, there is no doubt that the bright line spectrum in this case is the result of recombinations of hydrogen atoms. Hence the conclusion that U Gem stars, at least in epochs of constant magnitude, are surrounded by highly ionized envelopes. Since the emission of the envelope is very strong (in the Balmer continuum its intensity is higher than that of the star itself), its density should be sufficiently high.

The observed shifts of the bright and absorption lines in the spectrum of SS Cyg led Joy to the conclusion that this star is a spectroscopic binary. The excitation of the bright lines is attributed to a B-type star /16/. However, the available spectroscopic observations of SS Cyg do not reveal any signs of a hot star in the system (apart the obvious presence of bright lines). The continuum emission observed in the visible spectrum together with the continuous spectrum of a dG5 star should be the result of recombinations of hydrogen atoms in the envelope. Its color temperature is low. The second component of SS Cyg therefore greatly differs in its size and spectral energy distribution from a normal B-type star. The origin of the recombination spectrum of the star at the minimum therefore remains uncertain.

The emission of the envelopes of U Gem stars is closely related to the origin of their flares. The flare may be the result of a sudden increase in the temperature of the luminous surface or the result of an increase in surface area: it may also be the combined product of these two factors. According to present-day notions of nova explosion, the expansion of the

luminous surface makes a substantial contribution to the increase of the apparent brightness to the maximum. For U Gem stars, however, we have no data pointing to a rapid expansion of the envelopes. Conversely, it seems that the radius of the flaring star in the SS Cyg system (dG5, if its radiation is Planckian) is the same at the maximum and between the flares /17/, so that the flare is entirely attributed to the increase of star's temperature from 5000 to 15,000°.

This explains the variation of the hydrogen line profiles in the spectrum of SS Cyg on the rise and descending branches. The absolute intensity of the bright lines is apparently a function of the intensity of the absorption lines. As the radiation of a star with absorption lines in its spectrum increases in intensity, the bright lines are superimposed on the absorption lines, which thus appear weaker and eventually disappear. When the brightness of the star declines after the flare, the bright lines become progressively more prominent against the background of absorption lines.

It is not certain whether or not the radiation of the flaring component of SS Cyg is Planckian. The conclusion that the flare is associated with a sudden increase in temperature is thus tentative. This assumption does not exhaust all the possible mechanisms of brightness increase. As we have already noted, the radiation of some stars (T Tau, UV Cet, and others) according to Ambartsumyan is nonthermal during the flare. This factor may possibly operate in the case of U Gem stars as well. Z Cam stars are of particular significance in this respect. The light variation of these stars (Figure 103) follows the same trend as the light curve of U Gem stars over lengthy periods, but occasionally, after a $3-3^m.5$ flare the star drops 1^m and remains in this brighter-than-normal state for some time. After that the cyclic light variation is resumed. The spectra of Z Cam stars have been poorly studies. The scanty data available point to a likeness in the spectra of these stars with U Gem and sometimes T Tau stars /18/. This possibly indicates that Z Cam stars occupy an intermediate position between U Gem and T Tau stars, and all the three types are probably linked.

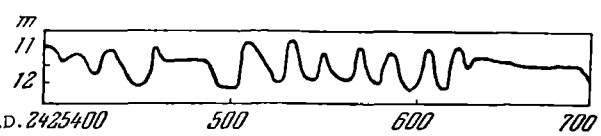


FIGURE 103

§53. R Coronae Borealis stars. Most of the time R CrB stars retain a constant brightness with slight fluctuations of $0.2-0^m.3$. Occasionally their brightness rapidly falls by $5-8^m$, and then gradually returns to its normal value (Figure 104). The luminosities of these stars at the time of constant brightness correspond to the luminosities of giants, and at the minimum R CrB stars are dwarfs. A considerable proportion of the radiation at the maximum is probably of nonthermal origin /2/.

Only few stars of this type are known, and almost all of them are so weak that their spectra (especially at the minimum) have been hardly

studied. The only exception is R CrB itself. This star has been studied in considerable detail in periods of constant brightness /19/ and at the minima /20/.

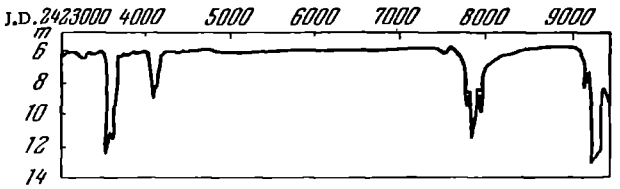


FIGURE 104

At periods of constant brightness the spectrum of R CrB is late F or G0 with the characteristic c, indicating that the absorption lines are deep and sharp. This shape of absorption lines is typical of giants. However, the intensity of a number of lines in the spectrum of R CrB differs from the intensity of the corresponding lines in the spectra of giants of the same type. In particular this is applicable to the exceptionally high intensity of the carbon lines and the anomalously low intensity of the hydrogen lines.

Apart from carbon lines, the lines of ionized metals and rare earth elements are also of high intensity. The average line shift corresponds to velocities of + 25 km/sec, which vary by a few km/sec in either direction. The energy distribution in the visible spectrum corresponds to 8000°, and in the violet spectrum it is 1000° higher.

At the beginning of the declining branch, the spectrum R CrB shows no changes, until the brightness has dropped by more than 4^m. Then emission lines appear at the center of the H and K absorption lines (ionized calcium, Ca II), and the absorption lines become diffuse and shallow, as if veiled. Further decrease in brightness leads to the appearance of bright lines of ionized and neutral metals. The width of the H and K emission lines increases to 300 km/sec and sometimes more. Forbidden emission lines [O II] are also observed in the spectrum at the same time. The emission Balmer lines, on the other hand, are absent.

The chemical composition of the atmosphere of R CrB was determined from the relative intensities of the absorption lines (assuming a state of thermodynamic equilibrium) /19/. The most abundant element in the atmosphere of this star was found to be carbon (69%), the content of hydrogen was much lower (27%). Note that not all the stars with light curves similar to that of R CrB are "carbon" stars. For example, XX Sgr is distinguished by anomalously high intensity of iron lines (the hydrogen lines being relatively weak), and is thus classified as an "iron" star.

To explain the peculiar light curve of R CrB, it was suggested that this star is subjected to flares, not unlike the nova outbursts. Since the atmosphere of the star is exceedingly rich in carbon, the shell produced by the flare is also rich in carbon and therefore opaque. Initially it absorbs most of the radiation emitted by the star, and then the brightness gradually increases as the shell is dissipated.

There are, however, a number of observational facts which contradict this hypothesis:

1) the intensity of the C I and C II absorption lines does not change when the brightness decreases;

2) the spectrum of the star remains unaffected as the brightness drops by a factor of up to 40;

3) not all the stars with R CrB-type light curves have carbon-rich atmospheres.

In the light of these objections, the hypothesis assigning the fall in brightness of R CrB to the absorption of stellar radiation in a carbon envelope proves inconsistent.

The following important point, first noted by Ambartsumyan /2/, is apparently of considerable significance for understanding the exact nature of these stars. The light curves of some R CrB stars are very similar to the light curves of T Tau stars. For some time, T Tau itself was actually classified as an R CrB star. The light variation of T Tau stars, in Ambartsumyan's opinion, is a unique phenomenon: it involves the conversion of the stellar interior energy into other forms of energy, partly radiation, in the outermost strata of the star. The same factor probably plays an important role in the case of R CrB stars.

§ 54. **RW Aurigae and T Tauri stars.** The different types of variable stars considered in the previous sections had their own characteristic light curve, which was either periodic or not. There are, however, some low-luminosity stars whose light curves are neither periodic nor of constant shape. These are the RW Aur stars /21/. The light variation amplitude of a given star may lie anywhere between $0^m.1$ and 3^m in different epochs, and the rate of light variation may change from 1^m in 1 hour to $0^m.1$ in 1 day. Periods of enhanced activity are sometimes followed by long quiescent periods. The light curve of one of the typical RW Aur stars — RR Tau — is shown in Figure 105.

FIGURE 105

The spectra of RW Aur are not uniform. Stars ranging between spectral types B and M are observed. G-type stars predominate. The spectra are often variable.

RW Aur stars include a separate group of stars whose spectra contain bright lines, not unlike the lines of the solar chromosphere (H, Fe I, Fe II, Ca I, Ca II, and other elements). The absorption spectrum (not always visible) of these stars corresponds to F8—M2 /23/. They are called T Tau stars.

The spatial distribution of T Tau stars is highly peculiar. Most of the known stars of this type are grouped in fairly small parts of the sky. The

same areas also show increased numbers of other RW Aur stars. Ambartsumyan, who was the first to call attention to this point, concluded that these aggregates of comparatively rare variable stars cannot be attributed to chance factors: stars found in these aggregates should therefore have common origin. Ambartsumyan called these groups T associations. Stellar associations, according to Ambartsumyan, must have formed fairly recently, since they are inherently unstable and tend to decay under the attraction of other stars. Since the associations are still observed intact, the age of the constituent stars is fairly low. According to Ambartsumyan's estimate, it is of the order of 10^7 years.

T Tau stars are thus of recent origin compared to most stars in the Galaxy (whose ages range between 10^9-10^{10} years) and therefore of considerable interest in astrophysics and cosmogony.

The principal spectral features /22, 23/ are the same for all T Tau stars, but their spectra also have some individual features. A common property of their spectra is the presence of bright lines. The strongest bright lines are the H and K lines (Ca II). The hydrogen Balmer lines also reach high intensity. The spectra are rich in Fe I and Fe II bright lines. T Tau spectra show emission lines of He I (and weak He II), and the forbidden line [S II] λ 4068.2 Å. On the whole, the emission spectrum is not unlike the spectrum of the solar chromosphere at heights above 1000 km, but the bright lines in T Tau spectra are much wider than in the chromosphere, reaching 4—5 Å.

The continuous spectrum of T Tau stars shows a superimposed continuum which veils the absorption lines. The continuum emission increases at shorter wavelengths, and the color of T Tau stars is thus bluer than that corresponding to its spectral type. The extremely rapid increase of intensity in the ultraviolet spectrum ($\lambda\lambda$ 3850—3650 Å) observed in two T Tau stars (NX Mon and VY Ori) /24/ and later in another star of this type, AG Dra /25/, is related to the same continuous emission. Another factor which may explain the marked increase in the ultraviolet intensity of these stars may be the merging of the Balmer emission lines which condense at the Balmer limit /24/.

The absorption spectra of various T Tau stars are nonuniform. Sometimes the absorption lines are completely obliterated by the continuous emission. If the absorption lines are prominent, they are either superimposed on bright lines or else they appear near the red wing of a bright line. The bright lines as a rule show a greater shortward shift than the absorption lines. This signifies that the atmospheric layers in which the bright lines originate move away from the star faster than the layers where the absorption lines form.

All T Tau stars have variable spectra. However, despite the intensive studies of these stars, the exact variation of the spectra have not been established. It only seems that the continuum emission and the line emission generally increases in intensity toward the maximum. There are also indications that the color temperature increases with the increase in brightness.

T Tau stars are often associated with diffuse nebulae, and in some cases these nebulae are also variable. Stars of this type very often appear in conjunction with variable comet-like nebulae. T Tau itself is situated in a

variable nebulae with angular diameter of about 10—12". The spectrum of the nebula shows high-intensity bright lines [O II] and [S II] and hydrogen Balmer lines. The star is also surrounded by a variable shell 4" in diameter, not unlike a semicircular prominence in shape. R CrA is the nucleus of a variable nebula. R Mon, RY Tau, and many other stars of this type are similarly situated. The variations of the nebula often do not show any apparent correlation with the variations of the star. The association between the star and the nebula in these cases, however, clearly cannot be regarded as accidental.

T Tau stars are dwarfs. The absolute magnitude of T Tau itself is close to 5^m, and the distances of the known stars of this type from the Sun are about 100—150 pc.

In some cases T Tau (or RW Aur) stars have close visual companions, which are dwarfs of later spectral types with bright lines. For example, the companion of T Tau, DD Tau, is a star of the same type, the companion of UX Tau is a dM2e star. Seeing that T Tau stars are of recent origin, we conclude that they formed as binaries from the start, and not through collision and capture. The great cosmogonic significance of this conclusion is obvious.

T Tau stars also include the so-called Herbig—Haro objects, which are weak stars situated inside small nebulae, with spectra similar to the spectrum of the nebula around T Tau /26/.

Some of the T Tau stars in the Orion Nebula discovered by Haro show extremely rapid variations of brightness. A flare takes only a few tens of minutes /27/.

Despite the considerable differences between the individual T Tau stars, they all show continuous emission, either all the time or intermittently during the flares. Proceeding from this fact and comparing it with the similar emission in a number of RW Aur stars and stars of other types (in particular, UV Cet stars, see below), Ambartsumyan came to the conclusion that the flares of T Tau stars involve the release of stellar energy from the interior by some hitherto unknown mechanism. Part of the released energy is converted to heat radiation and other forms of energy, and the remaining part is radiated as continuous emission (nonthermal radiation). In the case of T Tau stars, the energy is probably released at a great optical depth (below the photosphere) and the temperature of the star therefore increases during the flare /2, 28/.

The spectra of RW Aur stars are nonuniform. Some stars, in particular, RW Aur itself, have spectra which are similar to the spectrum of a T Tau star, whereas in others this similarity is hardly noticeable or nonexistent. The spectrum variations of RW Aur are similar to those of T Tau. For example, as the brightness falls from 10.7 to $11^m.3$, the color temperature (between $\lambda\lambda$ 4000—6500 Å) decreases from 7300 to 4300° /29/.

The luminosity of RW Aur stars is often close to the luminosity of the main-sequence stars of the corresponding spectral type.

The differences in spectra, luminosities, and other characteristics of RW Aur stars are the result of the application of photometric classification methods, i.e., the shape of the light curve and the magnitude, ignoring the other features of the stars. A more detailed classification shows that RW Aur stars do not constitute a homogeneous group, and should in fact be subdivided into a number of independent types.

§55. UV Ceti stars. The development of observational techniques during the last decade made it possible to observe the spectra of numerous weak stars, mainly those of late spectral types. The spectra of numerous stars were found to contain bright Ca II lines and sometimes hydrogen lines. The percentage of stars with bright spectral lines among types M4—M8 is higher than among early-type stars. Since the presence of bright lines in the spectrum is often an indication of nonstationary behavior, at least some of the red dwarfs are expected to be variable stars. In 1948—1952, the light of some stars of this type (dMe) was found to vary /32/. The best known of these is UV Ceti (L 726—8), which is the prototype of the entire group.

UV Ceti stars are characterized by extremely rapid and irregular "flares": the brightness of the star increases more than tenfold in a matter of minutes (at a rate of $0^m.1$ in 1 sec, and faster) and then rapidly declines. The greater the flare amplitude Δm, the longer is the time Δt for the star to recover its original brightness. According to Oskanyan, Δt is an exponential function of Δm /33/. The entire flare takes about an hour. Sometimes the light variation is even faster. Thus, the brightness of UV Cet in 1952 increased by $4^m.8$ in 20 sec. Large flares (with amplitudes $> 1^m.5$) are interspersed with smaller flares. The average period between successive small flares is about 30 min. Relatively slow variations of brightness (at rates of about $0^m.001$ per sec) are also observed /33/. Photometrically, the flares of UV Ceti stars are reminiscent of chromospheric flares in the Sun. Haro pointed to a considerable similarity between the flares of UV Ceti stars and the flares of the fast T Tau stars in the Orion Nebula /27/.

The spectrum of the star undergoes substantial changes during the flare /34/. The absorption lines and the molecular absorption bands are veiled by high-intensity emission of unknown origin. This emission is particularly intense in the short-wave "blue" region of the spectrum. The intensity of the blue hydrogen lines also increases and bright He I and He II lines appear.

Double stars are often observed among the known dMe-type stars and, in particular, among UV Ceti stars. UV Cet itself is a binary. The spectral type of both components is dM6e. Their visual absolute magnitudes are $15^m.6$ and $16^m.1$, the luminosities being $5 \cdot 10^{-5} L_\odot$ and $3 \cdot 10^{-5} L_\odot$, respectively /35/. The flaring component is apparently the weaker of the two. The energy released in one flare is of the order of $10^{33} - 10^{34}$ erg.

The rate of light variation of these stars is very fast. Seeing that the star recovers its original brightness soon after the flare, Ambartsumyan /2/ concludes that the flares cannot be the result of an increase in surface temperature or ejection of an envelope by the star. Either mechanism would involve a much slower decrease of brightness after the flare. Ambartsumyan therefore links the flares of UV Ceti stars (like those of T Tau stars) with explosive processes which convert the stellar interior energy into nonthermal radiation. In UV Ceti stars, this conversion apparently takes place in the outer layers of the chromosphere.

REFERENCES TO CHAPTER IX

1. AMBARTSUMYAN, V.A.— Astron. Zh., 26:3. 1949.
2. AMBARTSUMYAN, V.A.— Bull. Byurakan Observ., 13. 1954.
3. PLASKETT, H.— Publ. Obs. Victoria, 4:119. 1928.
4. TCHENG, Mao Lin, and M. BLOCH.— Ann d'Ap., 17:6, 1954.
5. BLOCH, M. and TCHENG, Mao Lin.— Mem. Soc. Roy. Sci. Liege, 20:458. 1958.
6. STRUVE, O. and P. SWINGS.— ApJ, 93:356. 1941.
7. MERRILL, P.— ApJ, 99:15. 1944.
8. PAYNE-GAPOSCHKIN, C.— ApJ, 104:362. 1946.
9. SOBOLEV, V.V. Dvizhushchiesya obolochki zvezd (Moving Stellar Envelopes).— Leningrad. 1947.
10. GAUZIT.— Ann. Astrophys., 18:354. 1955.
11. PARENAGO, P.P. and B.V. KUKARKIN.— Peremennye Zvezdy, 4:245. 1934.
12. GRANT, G.— ApJ, 122:566. 1956.
13. ZUCKERMANN, M.— Mem. Soc. Roy. Sci. Liege, 20:237. 1958.
14. ELVEY, C. and H. BABCOCK.— ApJ, 97:412. 1943.
15. HINDERER, A.— AN, 277:193. 1949.
16. JOY, A.— ApJ, 124:317. 1956.
17. GORBATSKII, V.G.— Astron. Zh., 29:6. 1952.
18. JOY, A.— PASP, 44:385. 1932.
19. BERMAN, R.— ApJ, 81:369. 1935.
20. HERBIG, G.— ApJ, 110:143. 1949.
21. KHOLOPOV, P.N.— Astron. Zh., 27:233. 1950. Trudy IV Soveshchaniya po Kosmogonii, p.367. 1955.
22. JOY, A.— ApJ, 102:168. 1945.
23. HERBIG, G.— In: "Nestatsionarnye zvezdy," p.95, Erevan. 1957.
24. BÖHM, K.— Zs.f.Ap., 43:245. 1957.
25. ARAKELYAN, M.A. and N.L. IVANOVA.— Bull. Byurakan Observ., 24:19. 1958.
26. HARO, G.— In: "Nestatsionarnye zvezdy," p.125. Erevan. 1957.
27. HARO, G.— Bol. Observ. Tonanzintla, 14:3. 1956.
28. AMBARTSUMYAN, V.A.— Trudy IV Soveshchaniya po Kosmogonii, p.344. Moskva. 1955.
29. HERBIG, G.— PASP, 60:256. 1948.
30. JOY, A.— ApJ, 105:96, 1947.
31. JOY, A.— Astron. J., 55:68. 1950.
32. LIPPINCOTT, S.— ApJ, 115:582. 1952.
33. OSKANYAN, V.— In: "Nestatsionarnye zvezdy," p.17. Erevan. 1957.
34. JOY, A. and M. HUMASON.— PASP, 61:133. 1949.
35. Van de KAMP, P. and LIPPINCOTT,— PASP, 62:365. 1950.

B. A. VORONTSOV-VEL'YAMINOV

Chapter X

STARS WITH VARIABLE SPECTRA

§ 56. **Spectroscopic variables.** Spectroscopic variables, i.e., stars with variable spectra, are also characterized by light variation. These include cepheids, long-period and irregular variables, novae, and nova-like stars. Detailed examination of a spectroscopic variable often led to the discovery of fluctuations, however small, in its magnitude. In some cases the spectrum appears to vary without any changes in the magnitude of the star, but sooner or later light variation is also observed. The spectra of eclipsing binaries, which are in fact the resultant spectrum of the entire system, also show definite changes associated with axial rotation of stars, ejection of gases from the atmosphere, etc. In these cases the spectrum variation is mainly associated with the changes in the contribution of the individual components to the resultant spectrum in different phases of the eclipse. The spectroscopic variables have been described in the appropriate chapters of this Course.

In the present chapter we will discuss hot stars with changing bright lines in their spectra, stars surrounded by a tenuous absorbing envelope, and stars with variable magnetic field.

§ 57. **Stars with variable bright lines.** The spectra of O, B, and A stars (especially of B stars) show bright lines, predominantly of hydrogen. In most cases only H_α is bright; occasionally bright H_β and H_γ lines are observed, and seldom H_δ as well; in certain cases, however, all the emission lines of the Balmer series show in the spectrum. In numerous cases the bright lines or some of them would appear or disappear after a few years. They have been best studied in B-type stars, and it seems that Be stars are spectroscopic variables to a certain extent. Stars with numerous bright lines generally show light variation as well. Besides hydrogen lines, their spectra also show bright lines of He I, He II, Fe II, and occasionally the lines of other elements.

There are over 1000 stars with bright spectral lines among O, B, and A stars; their percentage increases from the late types to B2 and then again decreases.

At this stage it is hardly possible to distinguish between B stars with constant spectra and stars whose spectra occasionally show emission lines. It is furthermore impossible to differentiate between stars with variable emission spectra and stars with variable brightness. The best example of this difficulty is γ Cas (B0ep), which for several decades was

regarded as a typical spectroscopic variable, and then in 1936 suddenly flared increasing its brightness by 1.5 stellar magnitudes. Stars continually observed over prolonged periods had pure absorption spectra, which suddenly developed a bright H_α line showing on panchromatic plates or other emission lines.

The bright lines in the spectra of Be stars generally show a distinct broadening, which increases with increasing wavelength in accordance with the general requirements of the Doppler effect. This broadening, however, is associated with neither thermal motions nor turbulence, since these factors produce entirely different line profiles. The broadening of the emission lines is accompanied by broadening of the dark lines, and their profiles definitely point to rotational origin of broadening. This problem was studied in considerable detail by O. Struve, who also considered data on rapidly spinning spectroscopic binaries. The percentage of stars with bright lines among B stars was found to increase with increasing equatorial rotation velocity (projected onto the line of sight): the increase was from 3% for stars with rotation velocities less then 50 km/sec to 50% for stars with rotation velocities exceeding 150 km/sec. The percentage of fast spinning B stars, on the other hand, definitely decreases at velocities above 75 km/sec. As is known, O and B stars on the whole have the highest rotation velocities.

These data led Struve to the assumption that the bright lines originate in thin disks or rings which are detached from the fast spinning star under the action of centrifugal forces, which reaches its maximum on the equator. In the Roche's model, the B stars reach the stability limits at equatorial rotation velocities of 560 km/sec; substantial radiation pressure, however, may cause separation of gas on the equator even at lower rotation velocities. (When comparing the calculated and the observed rotation velocities, one should remember that the apparent spin is less than the true spin because the spin axis is mostly inclined to the line of sight at an angle less then 90°.)

The broadening is different for lines of different elements, which means that the corresponding elements have different velocities in the originating ring.

For example, bright Fe II lines are visible only when the hydrogen lines are very strong. It naturally seems that parts of the gas ring revolve about the star in accordance with Kepler's law. In this case the radius of the Mg II ring will be 1/5 of the radius of the hydrogen ring, provided the differences in the excitation and ionization conditions in the interior and the exterior parts of the ring do not affect the propagation of Mg II emission inside the ring. To judge from the lines excited in the ring, its temperature is invariably lower than the temperature of the star, and the conspicuous absence of forbidden lines indicates that the width of the ring is comparable with the star diameter, so that the stellar radiation reaching the ring is only slightly diluted.

There is nothing in the observations to suggest that the bright spectral lines originate in the ring and not in the overall extended atmosphere of B stars. The centrifugal force clearly makes a substantial contribution, and we can naturally expect separation of gas from the star in the equatorial plane; this, however, may produce an equatorial disk and not a discrete ring at a certain distance from the star. The observed appearance and disappearance of bright lines seems to suggest an inherent

instability of the ring or the disk, but unfortunately the actual mechanism of these spectral changes is far from being clear, since the rotational velocity of the star as a whole cannot change and it is not clear why the flow of gas from the star to the ring should stop, while its dispersion is continuous.

The observations of γ Cas can be interpreted as reflecting changes in the rotational velocity of the pulsating, variable-radius layer where the spectral lines originate. Then if the gas separates from the surface of the pulsating atmosphere, of variable rotational velocity, the density of the ring and the emission conditions will not remain constant. The emission of the ring, the disk, or the atmosphere as a whole is naturally a kind of fluorescence treated by Rosseland. In a low density gas with highly diluted radiation, an atom with three energy states will mainly experience transitions of the type $1 \rightarrow 3 \rightarrow 2 \rightarrow 1$, i.e., ionization followed by recombination with the electron subsequently cascading back to the ground state. This chain of transitions produces the emission Balmer lines, as in gaseous nebulae.

In parts, however, the atmosphere may prove to be opaque to Balmer and occasionally even to Lyman line emission. These factors, together with probable temperature fluctuations and the existence of convection and eruptive phenomena in the atmosphere, are responsible for various peculiar features in the line profiles, line intensities, etc. The theory of various related problems was developed in considerable detail by O. Struve, V. V. Sobolev, and V. G. Gorbatskii (see Chapter XV).

Note that the study of line profiles and line intensity anomalies in various Be stars gives rise to a wide range of highly interesting problems in the theory of dynamic structure of stellar atmospheres and the theory of the dynamics of various atmospheric processes. Topics of considerable significance are the expansion of stellar atmospheres, expansion velocities, the presence of velocity gradients, etc. The theory of these processes has been developed in fair detail, but its comparison with observations is often hindered by the great variety of factors and processes taking place in the real dynamic envelope of the star.

The continuous spectrum of Be stars also presents a serious problem which has not been solved yet. These stars are on the whole fairly distant, so that their light is greatly affected by selective absorption in the interstellar space. To take this factor into consideration, however, we need the absolute magnitude of the Be stars, while it is not clear at all whether or not the Be stars are brighter than normal stars of the same spectral subtypes. On the one hand, emission spectra clearly prevail among the supergiants with their extended atmospheres. On the other hand, there are definitely some B stars of normal or even subnormal luminosity whose emission spectra are associated with their fast spin. The existence of these two distinct groups is possibly responsible for the contradictory results.

A dense extended envelope may naturally emit on its own in the continuum. This continuous emission is superimposed on the emission spectrum of the star and thus distorts the energy distribution, making the star appear redder.

A semblance of the equatorial gas rings is probably observed in close eclipsing binaries, whose components are mostly A and late B stars.

Suitable examples are provided by SX Cas, RW Tau, β Lyr. In some cases the ring encircles only one of the stars, whereas in other cases it embraces the entire system. The emission lines, if at all visible between the eclipses, are split into doublets. One of the doublet components is evidently produced by the receding half of the ring, and the other component by the approaching half; the dark line in the middle is associated with absorption in the part of the ring which is projected onto the star and is moving across the line of sight. During the first phases of the eclipse, one of the two bright components disappears, and at the end of the eclipse the other component vanishes. The intensity ratio of the red to the violet component is variable (these two components are apparently also observed in in some solitary Be stars). Close binaries are subjected to unusually intense eruptive processes, not unlike solar prominences, whose characteristic velocities are sufficient to produce gas jets between the fast spinning components and around them.

The gas rings, disks, and jets around spinning stars are obviously of considerable cosmogonic significance: they possibly constitute the primordial matter which, interacting with the dust nebulae through which many of the Be stars pass, form planetary systems.

The mass and the radii of these gas rings are unfortunately unknown. Judging from the visibility of spectral lines during the eclipse, the ring radius is double the star diameter. The intensity of the hydrogen lines thus suggests that the minimum mass of the rings is approximately equal to the Earth's mass, if ionization of the gas is ignored. The true mass of the ring may be several orders of magnitude higher.

Stars with extended absorbing envelopes were discovered later than stars with extended emitting atmospheres.

A typical star with an extended envelope is Pleione in the Pleiades cluster, which occasionally develops a Be spectrum. Sometimes it shows a pure absorption spectrum with wide hydrogen lines which are indicatory of fast rotation. At other times extremely narrow and sharp nuclei appear against the background of these wide hydrogen lines. Other strong narrow lines are also observed, which apparently originate in the slowly spinning cooler tenuous envelope. The spectral variations in Pleione, 48 Lib, and other stars of this type are accompanied by hardly perceptible light variations. All these stars are apparently fast spinning. The development of the extended absorbing tenuous envelope and its transformation into an emitting disk or ring apparently constitute different stages of one evolutionary process which is characteristic of stars at the limit of rotational stability. The interrelation between different types of these envelopes, their formation, and mutual transformation are still not fully elucidated, although the considerable material on hand provides fruitful ground for detailed exploration of this problem.

§58. **Variable magnetic fields in stars.** Be stars are not the only stars with variable spectra whose light variation is insignificant (with amplitudes of a few hundredths of a stellar magnitude) or nil. Belopol'skii was the first to study these stars, and in 1913 he detected periodic variations in the spectrum of α^2 CVn, and A-type star. The results of Belopol'skii, his followers, and other spectroscopists point to the following. Most lines in the spectrum of this star do not change, and the other lines, whose intensity

is variable, fall into two groups, so that the maxima of the lines of one group coincide with the minima of the other group. The period of variation is 5.47 days. Subsequently light variation with an amplitude of up to one tenth of a stellar magnitude and with the same period was detected. The largest intensity variation is observed for the lines of europium and other rare earths — gadolinium and dysprosium, whereas the other group of variable lines comprises the lines of terbium and numerous metals. Near the minimum the variable lines are split into two components, whereas the constant lines also have a constant radial velocity.

Much later, periodic variations (with a period of 9.295 days) were observed in the intensity of europium and chromium lines in an A0p-type star, HD 124248 = BD−18°3789. The intensity of these lines also varies in counterphase. Some spectroscopic variables of spectral type A also show strong rare earth lines, whereas others have variable metal lines. The latter group of stars, however, should not be confused with the so-called "metallic stars" of type A. Two other spectroscopic variables of this type, 73 Dra (20 day period) and ι Cas (a period of about 1.5 days), show light variation with the same periods and an amplitude of a few hundredths of a stellar magnitude. Some of the A stars are thus characterized by numerous anomalies. The A stars discussed above apparently have a peculiar chemical composition, but the spectral variation could not be associated with changes in the excitation and the ionization of the atmosphere.

In 1948 Babcock discovered the Zeeman effect in the spectrum of a spectroscopic variable 78 Vir, and later in the spectra of similar stars γ Equ (F0p), α^2 CVn, HD 125248, and others. Some stars were found to possess a constant magnetic field in the atmosphere, the field strength exceeding 1000 gauss. The magnetic fields are investigated by measuring the polarization of light in the variable lines and in their splitting components. The field strength of γ Equ reaches 900 gauss, and is apparently constant.

In α^2 CVn both the field strength and the polarity change. The extreme values reach −4000 gauss, when the ionized europium lines have maximum intensity, and + 5000 gauss half a period later, when the Cr II line is at its maximum.

The magnetic field of HD 125248 varies from + 7600 gauss to −7600 gauss in phase with the intensity variation of Cr I and Eu II, Cr II lines.

These are apparently fast spinning stars with the pole almost facing the observer. The magnetic field is clearly associated with their spin.

The magnetic fields in question are apparently different from the local magnetic fields of sunspots; rather the entire star has a dipole magnetic field whose poles approximately coincide with the poles of the spin axis. (South and North magnetic polarities are equiprobable among these stars.) Europium and chromium concentrate in different layers, one prevailing in the polar regions and the other near the equator. The atmosphere immersed in a magnetic field is in a state of large-scale turbulence with eddy pulsations and prevalent horizontal motions. The magnetohydrodynamic waves considered in Alfvén's theory produce magnetic pulsations, which may cause reversal of magnetic polarity. These variations are superimposed on the almost constant magnetic field. Self-excitation phenomena are thus associated with the star's spin and convection.

REFERENCES TO CHAPTER X

1. VORONTSOV-VEL'YAMINOV, B. A. Gazovye tumannosti i novye zvezdy (Gaseous Nebulae and Novae). — AN SSSR. 1948.
2. STRUVE, O. Stellar Evolution. —Princeton Univ. Press. 1950.
3. SOBOLEV, V. V. Dvizhushchie obolochki zvezd (Moving Stellar Envelopes).
4. Nestatsionarynye zvezdy (Nonstationary Stars). — Erevan. 1957.
5. HERBIG, G. Non stable stars. — Cambridge, 1957.
6. Etoiles à raies d'émission, Liege. 1958.

Chapter XI

NOVAE, SUPERNOVAE AND NOVA-LIKE STARS

§ 59. **Novae and nova-like stars.** Variables known as novae include a great variety of stars. Even greater variety is observed among nova-like stars (some of which are occasionally classified as novae).

Typical novae are variable stars with a light variation amplitude of no less than 7—8 mag which are definitely supergiants at the maximum. The novae flare only once, and after the outburst they return to their initial brightness. The characteristic spectra of novae vary in close relation to the phase of the light curve.

Recurrent novae are similar to typical novae with regard to the variation of their magnitude and spectrum, but they have been observed to flare at least twice. The light variation amplitude is generally less than or possibly equal to the upper limit value for the amplitudes of typical novae. It thus seems probable that typical novae also flare recurrently, although at longer intervals. Indeed, some of the novae originally regarded as typical later revealed their recurrent nature.

Supernovae are flaring stars not unlike the typical novae, but their maximum luminosity is much higher than that of the most luminous supergiants and novae ($M = -11^m$ and higher, up to -18^m). Their light variation amplitude is uncertain, but it should be really tremendous. Some of the extraordinary stars observed prior to the invention of the telescope are currently regarded as galactic supernovae. In the telescopic era supernovae have been observed only in other galaxies, where they become visible near the maximum, when their brightness is comparable with the integrated brightness of the host galaxy. The spectra of supernovae and their variation have not been conclusively interpreted, but they are definitely different from the spectra of novae and recurrent novae. It is not clear whether or not the supernovae constitute a direct continuation of the brightest typical novae. They may fall in an independent category which is consistently brighter at least by one order of magnitude than the novae.

Nova-like stars are variables which, irrespective of their spectrum, reveal occasional flares of any (possibly small) amplitude. Ordinary stars with spectra showing some of the bright lines or bands observed in one of the nova phases are also sometimes classified as nova-like stars. This term, however, should apparently be reserved for very hot medium- and low-luminosity stars with their spectra showing some of the features characteristic of typical novae in one of their stages of development.

Moderately hot variables with spectra showing a remote resemblance to the spectra of novae, such as U Gem and Z Cam stars, are better not classified as nova-like, since their physical nature is apparently different.

Late-type stars with bright spectral lines characteristic of the nova-like population and with irregular light variation are apparently all binaries comprising a cold giant and a hot nova-like star.

P Cyg stars are sometimes called permanent novae: although no light variation has been observed, their luminosity is exceedingly high and the spectra show bright lines with violet absorption satellites, as in the spectra of novae soon after the maximum. In fact, however, these hot stars are apparently closer to γ Cas spectroscopic variables than to novae. These seem to be stable stars of almost constant brightness with highly extended atmosphere. Deneb, for instance, cannot be regarded as a nova-like star or a permanent nova.

Some of the low-luminosity hot stars with nova-like spectra (either variable or stable) are apparently remnants of earlier nova explosions.

§ 60. Recurrent novae. The first recurrent nova was observed in 1902, when T Pyx flared up again, after an earlier outburst in 1890. It recurrently flared also in 1920 and 1944. However, this was recognized as a significant phenomenon only after the recurrent outburst of RS Oph in 1933; this star had been previously recorded as a nova in 1898. Kukarkin and Parenago found that cyclic variables showed a statistical increase in the length of the average cycle with increasing amplitude; they further established that the two recurrent novae, T Pyx and RS Oph, also satisfied this relation. Their extrapolation led to the conclusion that T CrB, which flared in 1866 with an amplitude of 7 mag, should explode again after 60—100 years. This prediction was brilliantly confirmed when T CrB indeed flared again in 1946.

The known recurrent novae are listed in Table 1.

TABLE 1

Star	Nova	Interval in years	Amplitude, m
N Ori	1677, 1750, 1892	83, 142	>5
T Pyx	1890, 1902, 1920, 1944	12, 18, 24	6.5
U Sco	1863, 1906, 1935	43, 29	>8
RS Oph	1898, 1933, 1958	35, 25	8:
T CrB	1866, 1946	80	11:
N Sge	1913, 1946	33	8.0
N Sgr	1901?, 1919	18?	>7

No uniform notation has been developed for novae: they are designated both as variables (Q Cyg) and as novae (N Cyg 1872, N Cyg No. 2), prefixed with N and followed by the year of occurrence or by the chronological number in the particular constellation.

The relation between the flare amplitude A and the mean recurrence interval P is apparently only statistical and highly approximate: the true amplitude of T CrB, which is apparently a weak companion of a 9 mag star, is about 11 mag, and not 7 as assumed originally. On the other hand, numerous novae with amplitudes of 7—8 mag have not recurred for several

decades. For an average amplitude of 11^m, extrapolation gives a recurrence period of 1000—3000 years. Some authors, however, disagree with this generalization of a relation obtained for recurrent novae to typical novae.

The light curves of recurrent novae are exactly identical or only slightly different from one flare to another. This points to fundamental identity of the relevant processes.

The magnitude and the spectrum invariably resume their initial state between the recurrent flares. They may be either constant or variable. The spectra are either continuous or with bright lines, and in the course of light variation the intensity of the bright lines varies between wide limits and they may completely disappear. A common feature of all these spectra are the indications of extremely high temperature between the flares. Sometimes, however, the high-temperature spectra with bright lines are also reminiscent of G—M spectra. This apparently shows that we are dealing with a double star comprising a cold giant and a hot high-luminosity companion. Such pairs have been observed in o Cet, R Aqr, Antares, VV Cep, and other systems. Sobolev and Menzel are of the opinion that a composite spectrum of this type may also be produced by a solitary hot giant surrounded by an extended cold atmosphere. The binary nature of T CrB, however, was observed visually; furthermore, when the star is near the maximum all traces of the cold spectrum are obliterated. In the phase of relative stability it is the spectrum of the hot component that sometimes disappears. All this may be due to a slight variability of the cold star, combined with strong variation of the hot star. After a long period of uncertainty as to the exact nature of T CrB, the orbital motion of the two components was determined from the spectrum and the mass of the hot component (the flaring star proper) was shown (1958) to be close to the observed lower-limit value of $2.6\,M_\odot$.

In symbiotic spectra, the contribution from the hot recurrent nova comprises the entire ultraviolet region, sometimes with H, He, and Ca^+ absorption lines, and bright emission lines of hydrogen and multiply ionized atoms in other parts of the spectrum (these emission lines are either of stellar origin or are characteristic of gaseous emission nebulae). The energy distribution in the continuous spectra of recurrent novae is identical to that in the spectra of very hot stars.

Proceeding from the existence of recurrent novae, whose spectra between the flares are identical to the spectra of typical novae soon after the maximum, and seeing that their spectra are invariably those of very hot stars of varying luminosities, Vorontsov-Vel'yaminov argued convincingly that the Sun cannot erupt as a nova. The recurrence hypothesis of typical novae as such does not remove this danger, since the Sun may be just approaching the recurrent nova stage, especially as its luminosity is close to the a v e r a g e luminosity of novae between flares. However, seeing that the dispersion of luminosities about the mean is extremely large and that the physical nature of the Sun is entirely different from that of novae, we conclude that the Sun w'll not flare as a nova.

The light curves of recurrent novae are characterized by a steep rise branch followed by a more gradual decline, which is nevertheless faster than in typical novae. Another remarkable effect was observed for T CrB:

after an extremely fast fall in brightness (7 mag in 20 days, back almost to its normal magnitude), a slow secondary rise began after 2 months, the brightness increasing by 1.5 mag in two months. The increased brightness persisted for about two months, and then gradually declined back to its normal value. In 1938, 8 years before the large flare, T CrB slowly increased in brightness by 1 mag and soon after that its spectrum showed strong bright lines of neutral and multiply ionized elements, as well as nebular lines.

It is particularly significant that the spectra of some nova-like stars (and their spectra only) occasionally show bright lines typical of the solar corona; these are the forbidden lines of multiply ionized calcium and iron.

§ 61. Light curves of novae. There are many different types of light curves, but in the great majority of cases the brightness increases very steeply in 2—3 days and then declines gradually over a period of several years (Figure 106). The maxima are either sharp (I) or rounded (II), followed by a smooth decline (S) or a decline branch with oscillations (O). In terms of the rate of decline, the novae are divided into fast, moderate, slow, and very slow. RT Ser and η Car were exceptionally slow novae. The brightness of RT Ser continued increasing for about 10 years, and its flat maximum extended over 20 years. η Car fluctuated from 3^m to 1^m over some 150 years and then dropped to 8^m fairly rapidly. MacLaughlin established the following common features for all the light curves (Figure 107): 1) The initial increase in brightness is very fast, and the rate of increase slows down toward the maximum. 2) The initial part of the declining branch (the first 3 mag) is smooth or with secondary maxima. For faint (distant) stars only this section of the declining branch is observed. 3) Transitional stage, the brightness falling by another 3 mag. This part of the declining branch is either smooth (S), oscillating (O), or dips abruptly by many stellar magnitudes (D), and then gradually recovers a level some 6^m weaker than the maximum. 4) Final decline —fairly smooth, the light curve falling to the same level as before the eruption. Before and after the eruption, the stars are either stable or slightly variable (more frequently the latter). The irregular oscillations reach their maximum amplitude in N Per 1901 (2^m).

Among the hundred known light curves, no two are quite alike in shape and scale, but it is clear from the preceding that, despite the great variety, they have certain common regular features.

A comprehensive classification of light curves advanced by the present author includes a brief description of the curves: the shape of the maximum and the rate of decline over the first 3^m (expressed as the logarithm of the time in days, without decimal point after the characteristic), the type of the nova, the amplitude in stellar magnitudes, and an indication of constancy (C) or variability (v) at the minimum. The light curve of N Gem 1912 is briefly described as follows:

II O,157, O,11v,

which reads: a rounded maximum followed by oscillations, decline by 3^m during 37 days after the maximum followed by strong oscillations, amplitude of 11 stellar magnitudes, variable brightness at the minimum.

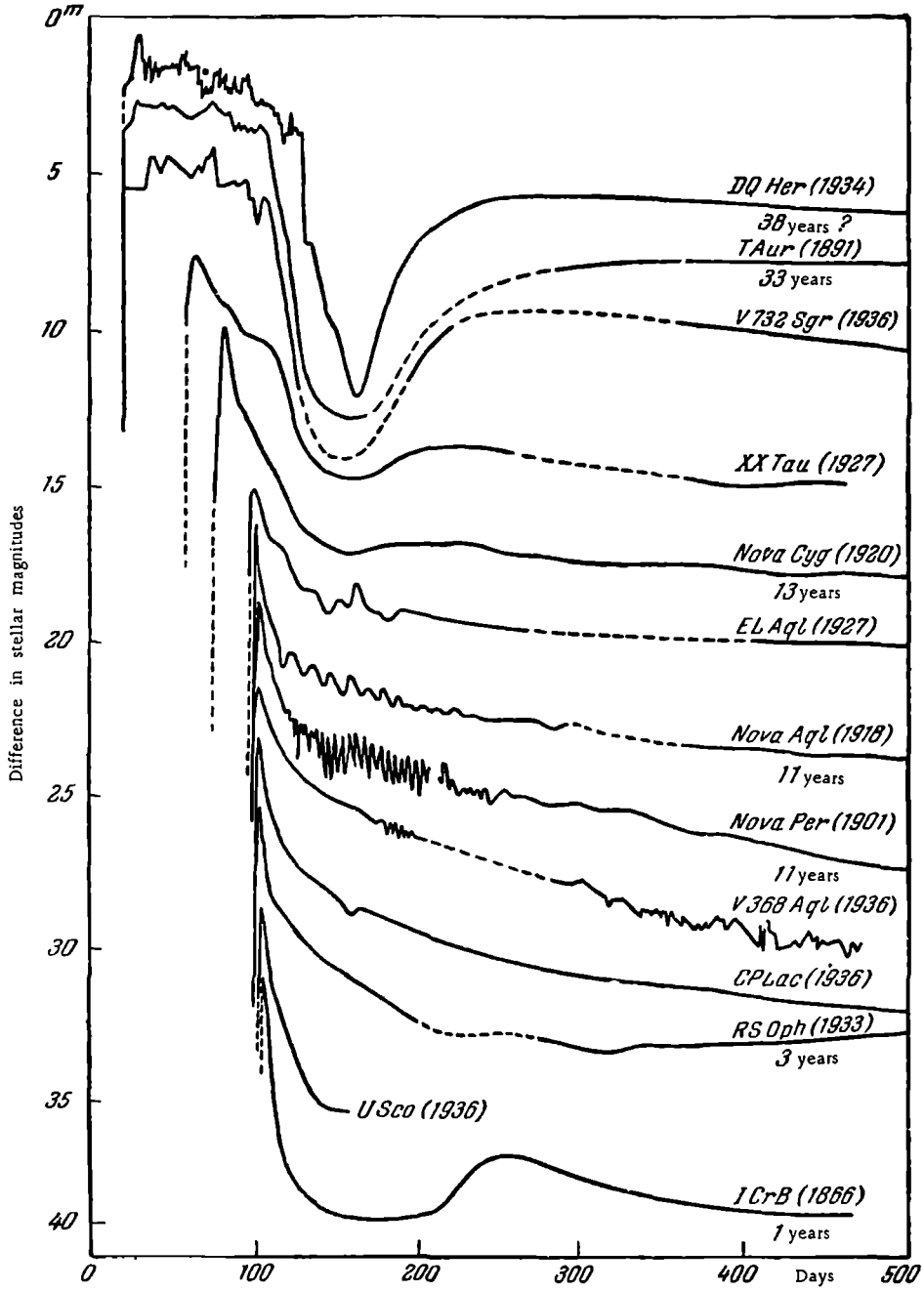

0m
5
10
15
20
25
30
35
40
Difference in stellar magnitudes
0
100
200
300
400
Days
500
DQ Her (1934)
38 years ?
T Aur (1891)
33 years
V 732 Sgr (1936)
XX Tau (1927)
Nova Cyg (1920)
13 years
EL Aql (1927)
Nova Aql (1918)
11 years
Nova Per (1901)
11 years
V 368 Aql (1936)
CP Lac (1936)
RS Oph (1933)
3 years
U Sco (1936)
T CrB (1866)
1 years

FIGURE 106

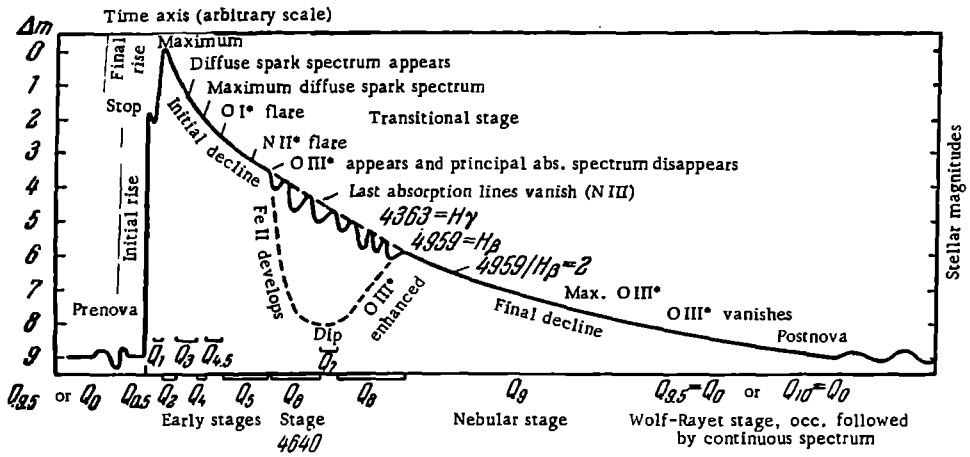


FIGURE 107

Light curves with smoothed ripples over the entire decline branch (Figure 108) are well fitted with the equation

$$m = m_0 + b \lg t,$$

where m_0 and b are measured three times on different dates. The logarithmic coefficients b characterize the rate of decline, and on the average the luminosity declines as

$$L = \frac{A}{t},$$

where A is a constant.

The overall amplitudes are from 6 to >17 stellar magnitudes; they are fairly uniformly distributed around the mean of some 11^m. Large amplitudes are naturally observed only for stars which reach high luminosities at the maximum. A number of factors in observations give too low amplitudes.

§62. Main spectral stages. The spectra of typical novae before the flare and in the very initial stages are unknown, with the sole exception of N Aql 1918, whose underexposed spectrum reveals it to be a white star. As we have noted, to ensure complete analogy with recurrent novae, the spectra of typical novae should be the same before and after the flare, i.e., these are spectra of hot stars. After the flare, as in recurrent novae, these are either continuous spectra or spectra with the characteristic bright lines and bands of Wolf—Rayet stars and with occasional traces of nebular lines (which show a tendency to disappear with time).

Soon after the maximum, when the first observations are generally made, the spectra of novae are those of ordinary A and F type supergiants with narrow lines shifted toward the violet by an amount which corresponds to approach velocities of a few tens or hundreds km/sec. N Her 1934 is the only nova which showed bright emission lines of hydrogen and metals

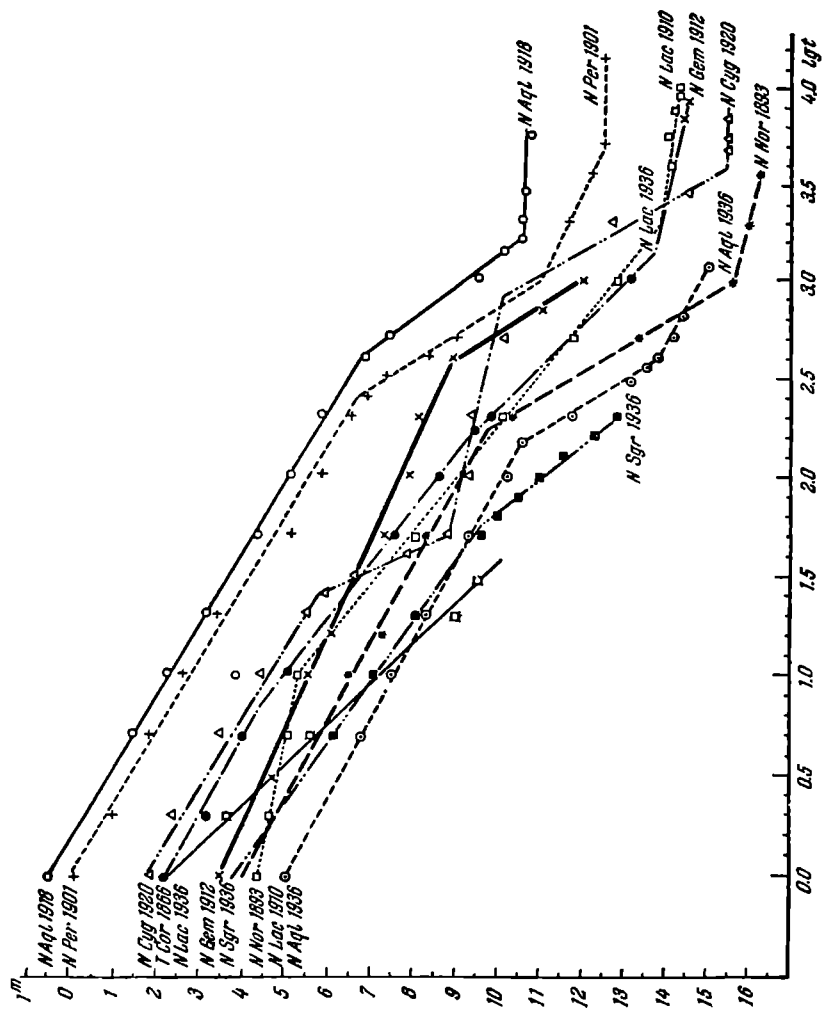

N Aql 1918
N Per 1901
N Cyg 1920
T Cor 1866
N Lac 1936
N Gem 1912
N Sgr 1936
N Vor 1883
N Lac 1910
N Aql 1936
N Aql 1936
N Per 1901
N Lac 1936
N Sgr 1936
N Gem 1912
N Cyg 1920
N Vor 1883
N Aql 1918
1^m
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 Lgt

FIGURE 108

a few days before the maximum, these lines disappearing at the maximum. The shift of the spectral lines of this nova also decreased toward the maximum.

The line shifts in this premaximum absorption spectrum should be interpreted as a result of the expansion of the star. The swelling of its luminous surface accounts for the increase in brightness and at the maximum the star is a true supergiant with radius exceeding hundredfold the solar radius and temperature corresponding to the observed spectral type.

At the maximum, the premaximum spectrum is rapidly replaced by the principal absorption spectrum, which shows a much greater shift (about 1000 km/sec) and is of a somewhat later type than the previous. Immediately after the maximum the principal spectrum develops wide bright bands in the original (unshifted) positions of its lines, which are flanked by greatly displaced absorption lines of the principal spectrum on the violet side.

This spectrum, with prevalence of hydrogen lines and a great number of iron and titanium lines, is the most characteristic spectrum of novae.

When the brightness of the nova has declined by about 1 stellar magnitude, the so-called diffuse spark spectrum develops, which comprises diffuse hydrogen and ionized metal lines shifted even farther toward the violet. It also has its own typical bright bands.

The diffuse spark spectrum coexists with the earlier principal spectrum and is gradually enhanced. At a certain stage it is reinforced by the "Orion" spectrum, which consists of the characteristic absorption lines (with emission bands) of the "Orion" stars, i.e., B-type stars. The Orion spectrum is also gradually enhanced.

Before the onset of the transitional stage of the light curve, the diffuse spark spectrum vanishes, the Orion spectrum is at its peak development, and a bright N III λ 4640 Å line appears. Then the Orion spectrum also vanishes, and the bright lines of gaseous nebulae appear. The intensity of these nebular lines increases. At the middle of the transitional stage, the last of the absorption lines of the principal spectrum disappear. At the end of the transitional stage the spectrum of a nova is reminiscent of the spectra of planetary nebulae; the only difference is that the stellar lines appear as wide bands and the continuous spectrum is not as strong as in nebulae. These spectra are characteristic of the nebular stage; they persist for a long time and are no less characteristic of novae than the earlier spectra. As the star approaches the stage of stable luminosity, and the spectrum is generally very weak, the nebular bands disappear and the spectrum becomes indistinguishable from that of a Wolf—Rayet star with He II bands and bands of multiply ionized atoms. This spectrum is sometimes retained permanently, showing secondary transient changes. In some stars the Wolf—Rayet bands gradually grow fainter, and the spectrum develops into a continuum, with bright emission in the ultraviolet.

If the brightness drops catastrophically in the transitional stage, the spectrum acquires numerous forbidden lines of ionized iron, which disappear at the beginning of the nebular stage.

Thus the succession of the spectra of novae also follows some regular fixed pattern, which is associated with the light variation phases, although

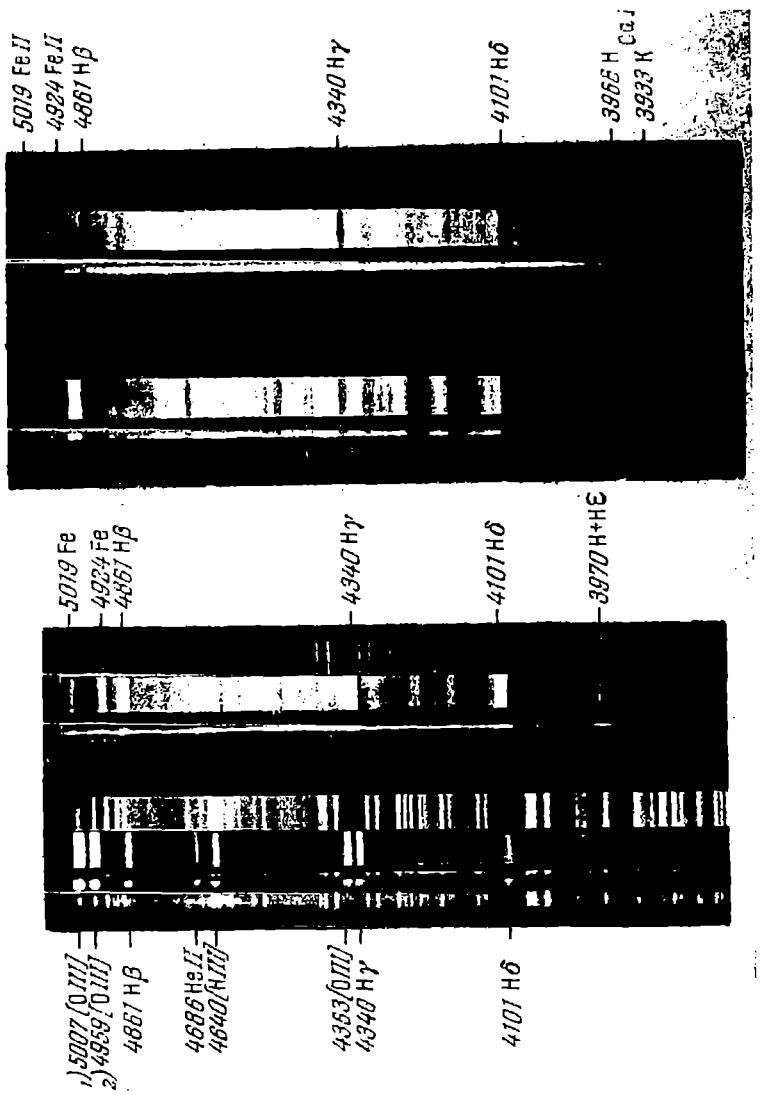

5019 Fe.II
4924 Fe.II
4861 Hβ
4340 Hγ
4101 Hδ
3965 H Ca.I
3933 K
5019 Fe
4924 Fe
4861 Hβ
4340 Hγ
4101 Hδ
3970 H+Hε
1)5007 [O III]
2)4959 [O III]
4861 Hβ
4686 He.II
4640 [N III]
4363 [O III]
4340 Hγ
4101 Hδ

FIGURE 109

the duration of these phases is determined by the rate of light variation and the light amplitude affects the length of the late spectral stages.

Figure 109 shows photographs of the spectrum of N Cyg 1920 in four principal stages of spectrum variation (from top to bottom: at the maximum, at the beginning of the declining branch, at the beginning of the nebular stage, and toward the end of the nebular stage).

The spectra of novae are designated Q, but in view of their great variety decimal subdivision is introduced, indicating the spectral phase.

In the course of secondary light fluctuations, the spectrum at the maxima develops into a "later" type and the continuous spectrum is enhanced. At the minima a reverse situation is observed.

Despite these common regular properties, the spectrum of each star has its own individual distinctive features, which are sometimes quite prominent. These individual features are reflected in band widths, line shifts, line structure, and relative line intensities. They are the outcome of certain characteristic differences in the process of nova eruption and in chemical composition.

Some stars, for example, show very strong neon, oxygen, or carbon bands, whereas in other stars the bands of these elements are weak. The nebular lines also reach various intensities. Forbidden lines of ionized iron or CN absorption bands are sometimes observed.

The structure of the bright bands is highly complex, especially in the nebular stage. They show individual maxima or split in phase with the semiperiodic light fluctuations of the nova.

The systems of absorption lines show a particularly complex structure in the epoch between the maximum and the nebular stage. A complex emission band is overlapped and limited on the violet side by several absorption lines, which sometimes belong to several components of the principal, diffuse spark, and Orion spectra, as well as to some other, less persistent spectra. Their intensities are all variable; the shift of these lines also changes systematically in one direction or other. The intensity sometimes fluctuates, and the number of dark components is different for different emission bands. Some stars show up to 9—10 systems of absorption lines. The shifts of the dark lines correspond to radial velocities from a few hundreds to 4000 km/sec.

The average shift of the principal spectrum lines linearly increases with the rate of decline of the nova. Correct interpretation of the spectra of novae and study of their changes requires extensive experience, great care, and of course a continuous series of high-quality spectrograms. The dispersion of the instrument is not particularly significant in this respect.

§63. **Nebulae illuminated and ejected by novae.** A filamentary, roughly circular nebula was discovered around N Per 1901 some six months after the eruption. The radius of the nebula increased at a rate of 11' annually, and its brightness declined, so that in 1902 the nebula became invisible. Its spectrum in 1902 was found to be identical with the spectrum of the star at the time of the maximum (1.5 years previously). This result, combined with considerations of the distance of N Per 1901 and the rate of growth of the nebula, showed that we were dealing in this case with a propagating light wave which illuminated the dust nebula as it advanced. The high-power light wave radiated by the nova at the maximum illuminated the originally invisible nebula which apparently surrounded the star.

Similar, though less prominent, effects are probably observed in connection with other novae also.

The assumption that the wide bright bands in the spectra of novae are associated with a transparent expanding gaseous nebula ejected by the erupting star at the maximum was confirmed by direct observations.

In 1916 a small nebula was discovered around N Per 1901, which is still expanding at a rate of about 1" annually. It had a typical gaseous spectrum with lines split along the middle. A transparent expanding nebula with a hollow interior will obviously account for this splitting, since at the apparent center the front hemisphere moves with maximum velocity toward the observer and the rear hemisphere recedes from the observer. Judging from the amount of line splitting, the nebula expands at a rate of 1200 km/sec; the expansion seems to have begun at the time of light maximum. Figure 110 shows some photographs of the nebula taken in 1917 and 1943.

FIGURE 110

Subsequently similar gaseous nebula were discovered virtually around all the bright, i.e., near, novae. Some 12 novae revealed this effect.

In 1928 splitting of N Pic 1925 was observed, and in 1935 splitting of N Her 1934; in both cases the two components receded from each other (Figure 111: a shows the star and the envelope in blue light, b is the envelope in the [O III] green light, c the envelope in [N II] light in July 1942, d-ditto in August 1942, e shows the structure of the [O III] lines). Subsequently these effects were identified with condensations in the nebulae ejected by the stars. Eventually these condensations dissipated. The bright nebular bands and their complex structure in the spectra of novae are associated with the structure of the ejected envelope and the orientation of the condensations relative to the line of sight. Inhomogeneity of the ejected envelopes is evidence of the explosive nature of nova eruption. Qualitatively the splitting explains some difficulties and transient features in spectra. Indeed, the absorption lines are produced by light being absorbed in different layers of the ejected envelope, which move with different velocities and have different temperatures and densities, that further change rapidly as the envelope expands.

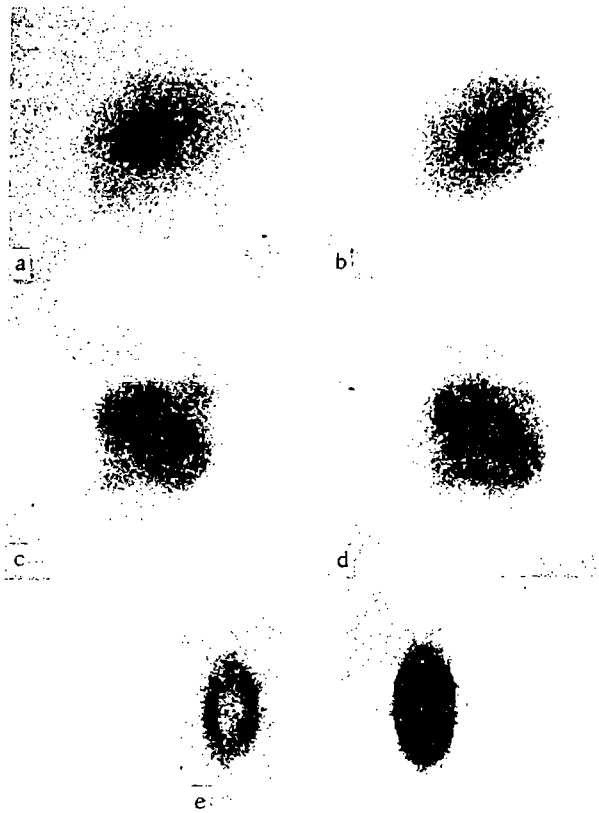


FIGURE 111

Direct observations of ejected nebulae have shown that the visible emission of the envelope in the nebular stage is considerably stronger than the star itself, the difference sometimes reaching a few stellar magnitudes. On the other hand, the brightness of the star measured by various techniques in fact corresponds to the resultant luminosity of the star and the envelope. The star itself accounts for the continuous spectrum, whereas the band emission is associated with the ejected envelope. If the so-called "bright band effect" is ignored, the observed brightness of a nova does not give a correct result for its radius and luminosity in different phases of evolution.

The nebular lines grow weaker in the combined spectrum after the nebula has dissipated appreciably. The nebula invariably expands at a uniform rate, much faster than the parabolic velocity. The formation of planetary nebulae following eruptions of novae is therefore definitely ruled out. Moreover, the ejected mass is three orders of magnitude less than the mass of the average planetary nebula.

The first masses of nova envelopes were determined by a variety of methods by Ambartsumyan and Kozyrev. They estimated the density of

the envelope (from its radius) at the time of the maximum, when the density begins falling. From that stage on, the continuum brightness decreases and the apparent magnitude of the star grows. An alternative method, applied by other astrophysicists, estimates the nebular mass from the intensity of the nebular lines. This method is based on Zanstra's theory of emission of planetary nebulae, whose masses are invariably determined in this way. A correction must be introduced for the concentration of ionized atoms whose lines are inaccessible to spectroscopic observations for some reason.

The various methods gave fairly consistent results. The masses of the ejected envelopes were found to range from 10^{-6} to 10^{-4} of the solar mass. The envelope mass gradually increases (after the shedding of the principal envelope, the star goes on ejecting stellar matter continuous), and the rate of increase is apparently higher the higher the maximum luminosity of the star.

While in the Wolf—Rayet stage, the nova apparently also loses some of its mass by continuous ejection of atoms from the surface, though at a slower rate.

§ 64. **Physical conditions in nebulae.** The emission of ejected envelopes seems to be partly due to photoelectric ionization of atoms followed by their recombination (as in gaseous nebulae) and partly to collisons with faster particles continuously ejected from the surface. The increase in the temperature of the exciting star and the decrease in the density of the expanding envelope gradually raise the ionization. When the dilution of radiation, which increases with distance from the star, reaches a certain level (corresponding to densities of about 10^{-19} g/cm^3), the envelope emits the characteristic forbidden lines of gaseous nebulae.

Unfortunately, it is not clear whether the envelope is detached from the nova at the beginning of the eruption or only at the maximum phase. At any rate, an envelope separating at the maximum should be much thicker and denser than the reversing layer of supergiants.

The electron density in the envelope drops from 10^9 cm^{-3} to about 10^5 cm^{-3} at the peak of the nebular stage; these values are close to the densities of planetary nebulae. The electron temperature of the envelopes is determined from the intensity ratio of the nebular lines, using the theory of emission of gaseous nebulae; it is found to range from 4000 to 9000°, remaining fairly constant with time. These figures are also close to the electron temperatures of nebulae. The likeness of processes and conditions, which continuously change in the envelope of each individual star, makes their study particularly valuable for the understanding of processes in nebulae, which expand at a much slower rate. However, the theory of the processes in rapidly expanding envelopes is much more complicated.

Sobolev's theory gives a good fit for the intensities of the bright lines. A much more difficult task was to obtain an adequate fit for the various emission band profiles predicted by the theory in the spectra of expanding transparent nebulae. In principle emission band profiles show whether the motion of atoms in the envelope is accelerated or decelerated, whether on receiving momentum the particles move by inertia or otherwise, and so on. Comparison of the different theories with observations is very difficult, however, since the profiles of the bright bands are partly distorted by superimposed absorption lines.

The various spectral features of expanding envelopes were treated in great detail by Mustel. In his opinion, the envelope is detached together with the reversing layer (whose thickness is comparable to the radius of the star at the maximum). The powerful luminous flux of the tremendous photosphere hitting the envelope from below imparts an additional momentum to the lowermost layers. The atoms ejected at a later stage with very high velocities from the surface of the star catch up with the envelope and sustain its emission by collisional excitation. The momenta transferred from these atoms to the envelope increase its velocity from the inside to the outside, and the premaximum spectrum is replaced by the principal spectrum with greater Doppler shift.

The diffuse spark spectrum is produced by high-velocity atoms ejected at a later stage from the exposed hot layers of the star. Atoms ejected from the still hotter interior layers are responsible for the Orion spectrum. The short-lived dark lines are produced by gas clouds intermittently ejected from the interior; these clouds rapidly dissipate or move aside so that they are no longer projected onto the star and stop absorbing the photospheric light.

Various complex features in the envelope spectra, however, still have no explanation. The identification and interpretation of the apparent features observed in the spectra on different days are possibly not always correct in view of their complexity and variability, so that the theory is sometimes concerned with nonexistent effects.

Radiation pressure is apparently not the only factor in the expanding atmosphere responsible for the separation of the envelope: explosive effects, possibly with the participation of electromagnetic forces, also make a certain contribution. The shedding of the envelope is analogous to a certain extent to the ejection of solar prominences, although this is a phenomenon on an entirely different scale. The structure of the envelope ejected by novae points to the existence of magnetic fields.

§65. Physical effects in the eruption of novae. While the processes which take place in the ejected envelope are clear in the main, the behavior of the erupting star has not been elucidated. Certain difficulties arise because of the absence of spectroscopic data for the rise branch of the light curve, except the phase immediately preceding the maximum.

At any rate, it is clear that a very hot star, similar to O-type stars before the eruption, will evolve to a cooler A or F supergiant at the maximum. Near the maximum the spectrum of the star is mainly continuous. The rise branch is therefore mainly associated with an expansion of the luminous photosphere, which also compensates the fall in temperature. Calculations show that prior to the eruption the star is hot with radius somewhat larger or much smaller than the radius of the Sun, whereas at the maximum the diameter of the star exceeds the solar diameter at least hundredfold.

The increase in brightness due to the expansion of the photosphere is confirmed by the shift of the absorption lines in the spectrum: this definitely shows that the front hemisphere of the star moves toward the observer.

Comparison of the linear expansion velocity obtained from line shifts with the rate of expansion calculated from the observed increase in brightness gives the radius of the star at different phases of the rise

branch. The results, however, are at variance with some other findings. On the rise branch of N Her 1934 the hydrogen and metal lines had different shifts, which decreased with time and became equal at the maximum (Figure 112).

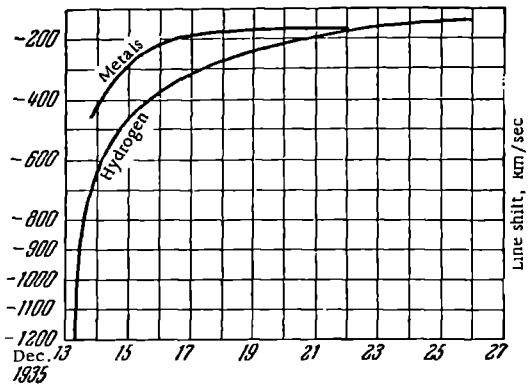


FIGURE 112

The theory of absorption line profiles in the spectrum of an expanding star leads to the conclusion that the lines should shift, expand, and become highly skewed. In fact, however, the skewness of the shifted lines is somewhat different from the theoretical prediction, probably because of interaction with nearby emission lines. The actual line width is much less than predicted.

These effects are generally explained as follows.

Long before the maximum, the photosphere becomes highly extended and the limb darkening is much more pronounced than for the Sun. The extent of this photosphere in fact determines the brightness of the star.

The reversing layer responsible for the absorption lines in the pre-maximum spectrum is also extended, and its expansion velocity may differ from that of the photosphere. The expansion of the luminous surface is therefore not quite equal to the expansion of the reversing layer which produces the line shift in the spectrum. Moreover, the reversing layer of N Her was stratified. The metals apparently began moving away from the center at an earlier stage, whereas hydrogen started later but with higher velocity. The hydrogen layer gradually caught up with the metals, the two components intermixed and continued moving with some common velocity up to the maximum. According to Mustel's calculations, the radius of the photosphere of N Her at the maximum reached 330 solar radii, and the radius of the reversing layer was 500 solar radii with a thickness of 170 solar radii. After the maximum the brightness fell and the envelope separated from the star.

So far no decision could be made between two competing hypotheses. According to one point of view, the star expands as one whole from the time of eruption to the maximum, and after the maximum it shrinks to its initial

radius. According to another point of view, the star as a whole does not change its radius. The outermost layers are detached at the very first phase of the eruption process; they form the expanding envelope which thereafter is observed as a growing nebula. The declining branch begins at the point where the expanding envelope becomes so thin and rarefied that its continuous emission is insignificant, and the exposed hot surface of the star shows through. This apparently explains the steady increase in temperature after the maximum.

The increase in temperature after the maximum is evident from measurements of the distribution of energy in the continuum; before the maximum the spectrophotometric temperature falls and it actually continues falling for some time after the maximum. The minimum temperature reaches 6000—7000° and under certain conditions CN absorption bands appear.

Further increase of temperature leads to the appearance of lines of multiply ionized elements in the spectrum. Application of the theory of emission of gaseous nebulae to the bright bands of the envelope spectrum gives the temperature of the exciting star. At the peak of the nebular stage, the temperature of the star reaches 100,000°.

During the secondary brightness fluctuation, the temperature falls at the maxima and increases in the minima. Belopol'skii's spectrograms of N Per 1901 give temperatures of 30,000° for the maxima of the transitional stage and 40,000° for the minima (allowing for the bright band effect mentioned in the preceding). The radius of the star was thus found to oscillate from 0.5 to 1.5 solar radii, increasing at the maximum. The electron density in the envelope changed accordingly. These findings show that the secondary maxima on the light curve in no way can be interpreted as flares against a background of generally diminishing brightness. The minima are conversely periods of significant deficiency in the radiation of energy by the star.

Figure 113 shows curves of variation of the apparent magnitude, the continuum intensity, and the shift of three systems of absorption lines for N Aql 1918.

Such oscillations are apparently the result of continuous ejection of gas from the star, which continues after the detachment of the envelope. When the ejected gas becomes transparent (e.g., following a decrease in the rate of ejection or acceleration of expansion), the observer sees radiation originating in deeper lying hotter layers of a smaller radius, and the apparent magnitude of the star increases. When the opacity of the expanding atmosphere is high, the brightness increases, but the effective temperature falls. Though the star on the whole may not contract, its effective radius after the eruption definitely decreases since after all the ejection of gases forming a dense dynamic atmosphere with continuum emission around the star stops eventually.

The masses of novae constitute another highly important topic. Binaries with known orbital motions were not observed among the novae for a long time, so that no direct determination of their masses was possible.

The absorption lines in the spectra of novae before and after the maximum reveal systematic variations, their shifts increasing and decreasing. The decrease of the line shift is regarded by some authors as the result of gravitational deceleration of the shell, caused by the attraction of the

central star. If this is indeed so, the results give only lower-bound masses, since the gravitational deceleration is partly offset by radiation pressure from the inside.

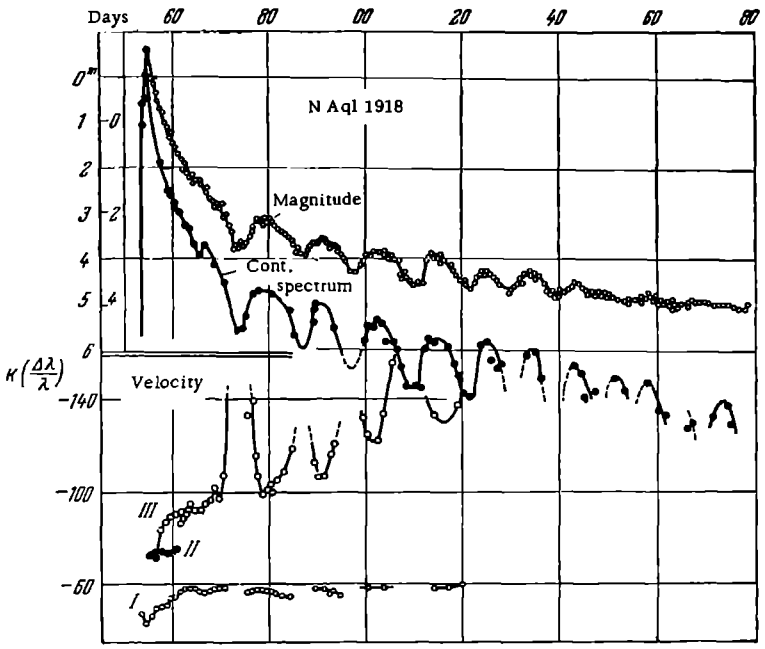


FIGURE 113

The masses calculated in this way were found to range between wide limits for different stars and also for the same star according as different lines were used. The masses of N Gem 1912 were found to fall between 150 and 1170 $M_\odot$. On the other hand, the various masses calculated for N Her generally do not exceed the mass of the Sun.

Walker (1945) made the astounding discovery that N Her 1934 was an eclipsing binary with an extremely short period of about 4.5 hours. Irregular light fluctuations superimposed on an Algol-type curve interfere with the exact determination of the orbital elements, the radius, the temperature, and the mass of each component from the light curve. It is clear, however, that the mass is at most equal to a few solar masses, and is probably much less than the mass of the Sun. Alongside with these irregular features, strictly periodic oscillations of variable amplitude (up to $0^m.1$) and a period of a mere 1.18 min were discovered. The companion of N Her 1934 is a cold dwarf. N Aur 1891 was also found to be an eclipsing variable with extremely short-period oscillations.

These discoveries further complicate our notions of the physical nature of novae. It is particularly intriguing to establish whether or not N Her 1934 was a binary before the eruption and which of the two components (if not both) actually erupted. One of the components is 3 mag weaker than

the other. The distance between them is a mere 300,000 km. Since among all the other stars even the masses of supergiants do not exceed 100 solar masses, and novae in the normal state are stars of average or even very low luminosity, most astrophysicists regard with considerable reservation the disproportionately large masses obtained from spectral line shifts.

The energy lost by the star is another important factor which has bearing on the physical changes occurring during the eruption. Detailed calculations were performed by Vorontsov-Vel'yaminov, who obtained a great diversity of values for different stars. No absolute figures could be obtained, however, since the masses are not known and it is uncertain whether the envelope is detached at the beginning of the eruption or at the maximum.

The energy loss is made up of three components: 1) the energy expended in separating the envelope of mass m from the star (its mass being M and radius r):

$$E_0 = G\frac{mM}{r};$$

2) the energy E_p expended in the expansion of the ejected shell of volume V at temperature T; this energy is a function of time,

$$E_p = \int p\,dV = \frac{R}{\mu} m \int T\frac{dV}{V},$$

where μ is the average molecular weight and R is the gas constant; 3) the kinetic energy E_k of the envelope; 3) the radiant energy E_r radiated by the star during the explosion, minus the energy radiated by the normal star in the same period.

Possibly a better characteristic than E_r or the absolute energy losses is provided by the ratio of these losses to the normal radiation of the star in a comparable period or in the period between two successive eruptions.

These ratios in fact characterize to what extent the eruption constituted a catastropic phenomenon and what its contribution to the overall energy balance of the star was.

If the envelope separates at the maximum, E_r is the principal loss component; it is virtually independent of the light variation amplitude, $E_r \approx 10^{45}-10^{46}$ erg for all novae. This is 10^4-10^6 times the annual energy radiated by the Sun. If, however, the shell is separated in the initial stages of eruption, E_0 is slightly greater than E_r.

On the whole the energy losses in both cases are less than the quantity of energy required to convert a normal star into a dense white dwarf. Energy considerations thus also show that eruption alone cannot involve radical changes in the interior stellar structure. The eruption does not alter substantially the general structure of the star and apparently only affects the outermost layers.

The ratio of the energy lost in the eruption to the energy radiated between successive eruptions is of the order of unity for typical novae (even if millennia pass between successive eruptions). In other words, the eruptions do not make an outstanding contribution to the total energy budget of the star. Let E be the total energy radiated by a normal star in a period required to radiate the entire explosive energy at the rate corresponding to the maximum of the light curve. Then the ratio E_r/E is

found to increase with amplitude, reaching about 50 for recurrent novae and sometimes exceeding 500,000 for typical novae. For very slow novae (η Car), E_t/E is as high as 10^7. This ratio is a measure of the disturbance in the energy balance of the star during the eruption. Its value provides a better description of the nova than the amplitude, since it allows for the individual shape of the light curve. The various hypotheses of nova eruptions naturally must explain all these data.

§ 66. Novae in other galaxies and their position on the Hertzsprung—Russell diagram. Spatial distribution. We have noted that at the maximum the novae are supergiants, but no exact data are available on their maximum luminosities. The novae are distant stars, their trigonometric parallaxes are unreliable, and their proper motions are poorly known.

Novae were repeatedly observed in other galaxies. In M31 (the Andromeda Nebula), 108 novae were recorded. Arp carried out a special observation project during 1.5 years and concluded that 26 novae should erupt annually in M31. Their light curves and spectra (observed only near the maximum) are identical to those of novae in our Galaxy. Their frequency increases toward the center of the Andromeda Nebula. The Magellanic Clouds contain very few novae, and only one nova has been so far observed in a galactic globular cluster. It is therefore not quite clear whether or not the novae are preferably associated with a certain star system. Observations in other galaxies at known distances provide an indication of the absolute magnitude at the maximum. For the novae in M31, Hubble found $M = -7^m.0$ on the old distance scale, which gives $-9^m.3$ on the new scale.

The determination of absolute magnitudes by various statistical methods (from the intensities of interstellar calcium lines, their shifts due to galactic rotation, etc.) gives the average value $M = -7^m.5$. Payne-Gaposchkin found $-7^m.6$ (1954).

One of the best methods for obtaining the individual parallaxes is by comparing the angular rate of expansion of the ejected envelope with its linear velocity (measured from the spectrum). Some authors, however, overestimate the value of this method: after all, there is considerable divergence in the various estimates of the angular diameters of nebulae and it is not clear what line-of-sight velocity is to be assigned to the given expansion velocity.

More detailed analyses have shown that the luminosity of a nova at the maximum is apparently related to the rate of decline soon after the maximum by a logarithmic formula of the form

$$M = -10.5 + 2.2 \lg t_3,$$

where t_3 is the number of days for the brightness of the nova to decline by 3^m from the maximum. For very fast novae $M = -8^m.2$, for moderate novae $-6^m.5$, and for the slowest novae (RT Ser) $M = -3^m.6$.

Given the individual absolute magnitudes of novae at the maximum (from the steepness of the rise branch and other data) and their amplitudes, we can find the absolute magnitudes before and after the eruption. The luminosities after the nova stage are found to range between wide limits, from $+1^m$ to $+9^m$ and fainter. The normal luminosities of novae are thus characterized by great diversity. This, and not their likeness to the Sun's

luminosity, is the typical feature of novae in the normal state. In terms of luminosity, the brightest of these stars are close to A-type main-sequence stars, and the faintest are reminiscent of white dwarfs. The tremendous range of normal luminosities of novae rules out the suggestion that the erupting stars constitute a physically homogeneous, unique group. Different evolutionary stages are present among these stars.

Since the dispersion of luminosities at the minimum is much greater than the spread at the maximum, recurrent novae are generally the brightest in the normal state, and stars with a small amplitude of the light curve which erupt less frequently are mainly white dwarfs. The temperature of novae in the normal state cannot be found with any precision because of fundamental measurement difficulties for stars with extended and stratified envelopes. At any rate, their temperatures are the highest among all known stars, approaching the similarly high temperatures of O—A stars, Wolf—Rayet stars, and nuclei of planetary nebulae.

Between successive eruptions, the novae on the whole form the so-called white-blue sequence in the Hertzsprung—Russell diagram (discovered by Vorontsov-Vel'yaminov in 1946). Together with the nuclei of planetary nebulae, which are characterized by a somewhat smaller dispersion of luminosities, they occupy a vertical strip on the HR diagram which begins in the field of white (more precisely blue, the hottest) dwarfs and does not quite reach the normal O—A stars. The gap between the two sequences is completely filled with Wolf—Rayet stars, which have a similarly large dispersion of luminosities. It is significant that in the HR diagram the gap between the stable hot giants and dwarfs is continuously filled by the least stable of stars, such as novae and possible Wolf—Rayet stars and nuclei of planetary nebulae. Greenstein's recent research has shown that other low-luminosity hot stars also exist, somewhat masking the white-blue sequence.

Knowledge of the abolute magnitudes of novae make it possible to study their distribution in the Galaxy. They show a pronounced concentration toward the galactic plane, and even more marked concentration toward the center of the Galaxy; on the whole the novae constitute an intermediate system between the disk and the halo components.

Their distribution is greatly reminiscent of the distribution of planetary nebulae; they are furthermore close to the nuclei of planetary nebulae in terms of their luminosity, dispersion of luminosity, and temperature. The natural question to ask at this point is whether the planetary nebulae are in fact gaseous envelopes ejected by novae. However, the masses of the planetary nebulae are 3—4 orders of magnitude greater than the masses of envelopes ejected by novae; moreover, the nova envelopes expand very rapidly and in no case did they decelerate to produce a stationary nebula around the ex-nova. Many of the nova envelopes are exceedingly small, so that they are invisible even through the largest telescopes despite their relative proximity to the Sun.

Some novae apparently erupted in the intergalacitc space, where as we know RR Lyrae stars, globular clusters, and possibly a multitude of other stars exist. Fragmentary observations of these novae, however, are insufficient to prove their existence.

The spatial distribution of novae in M31 is very clear. They occur all the way out to the visible edge of the galaxy, mainly in the direction of its major axis, showing a marked concentration toward the center.

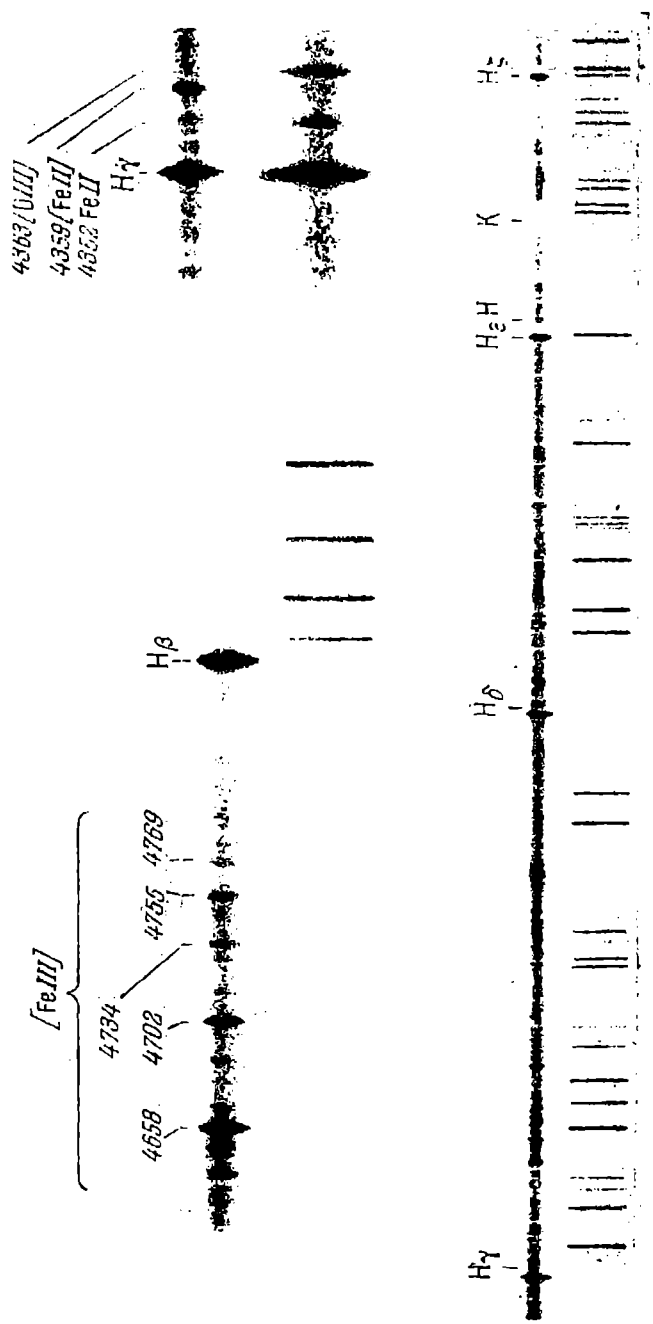

4363[OIII]
4359[Fe II]
4352 Fe II
Hγ
Hβ
[Fe III]
4734
4702
4658
4755
4769
Hγ
Hδ
HεH
K
Hζ

FIGURE 114

§67. **Nova-like stars.** In §59 we have mentioned the ambiguous definition of this class and the considerable uncertainty in the classification of variables and even some stable stars assigned to this category. In this section, we will not consider P Cyg stars, U Gem and Z Cam variable stars, and hot variables of the γ Cas type. Peculiar stars, such as T Tau stars, are not considered either. They were treated in Chapter IX.

The typical specimens of the "nova-like" group are hot stars whose spectrum and imperceptible or slight flucutations of magnitude are reminiscent of novae between eruptions. All these stars have high apparent magnitudes ($11-12^m$). Many of them are thus probably real novae which have not erupted so far in the brief epoch of our science.

Note that the spectra of some of these stars show typical features of G, K, and M stars. Here, as in the case of T CrB, we are probably dealing with binaries comprising a cold star and a hot nova-like companion.

The most typical nova-like stars are Z And ($8^m.9-12^m.8$), R Aqr ($7^m.5-12^m.0$), BF Cyg ($9^m.3-13^m.4$), CI Cyg ($10^m.5-13^m.2$), AX Per ($9^m.4-12^m.5$), RX Pyx ($11^m.1-14^m.1$), and others. Figure 114 shows the spectrum of BF Cyg.

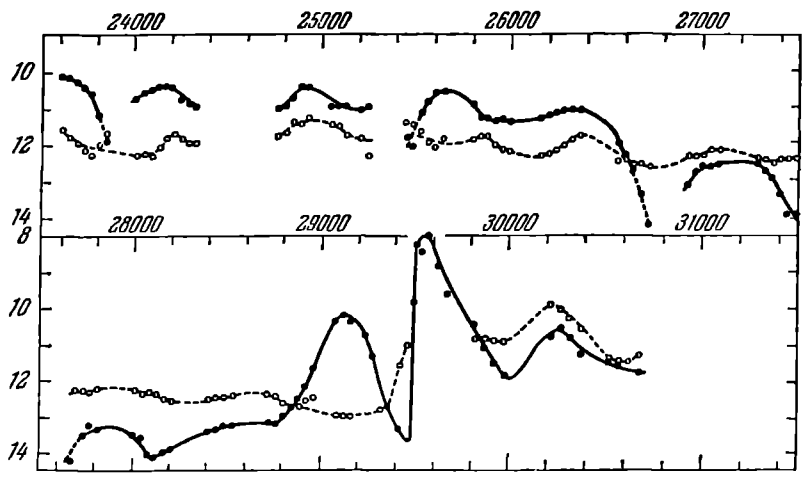


FIGURE 115

V Sge ($9^m.5-13^m.2$) is similar in all respects to these stars, but its spectrum does not contain a late-type component. The spectra of all these stars, with rare exceptions, contain wide emission bands of hydrogen, helium, ionized helium, nitrogen, and other elements on a continuous background, and also the forbidden lines characteristic of gaseous nebulae. Tiny nebulae are actually observed around Z And and R Aqr, and it is here that the nebular lines are excited. The relative intensities of the lines and the background are variable, as in novae between eruptions; the variations in spectrum are particularly pronounced when the light is variable and, as in novae near the maximum, the high-temperature continuous spectrum is enhanced and the cold spectrum grows fainter and may even disappear

altogether. This provides evidence of the binary constitution of these stars, where the hot, nova-like component is the more variable of the two. It is only in R Aqr that the cold component is a long-period red variable with an amplitude of up to 4 mag. Z And shows variations of radial velocity with the period of light variation (about 680 days). Some authors tried to resolve the observed light variation curves of Z And and R Aqr into the individual curves of the cold and the hot components. Figure 115 shows the light curve of Z And resolved into the curves of the red (circles) and the blue (dots) components.

Some of these stars near the maximum show P Cyg lines—wide bright lines with dark violet satellites shifted toward the violet as in the spectra of novae. The nebular lines in this phase disappear or grow weak, i.e., these spectra show a great resemblance to the evolution of typical nova spectra.

Z And stars provide a unique possibility of a simultaneous study of a great many emission lines, since the numerous emission lines in their spectra are all narrow and do not overlap. All these stars show considerable fluctuations of structure and radial velocity. The ejection of gas in these stars is not very powerful.

§68. **P Cygni stars.** As we have mentioned at the beginning, these stars cannot be considered as "permanent novae," especially in view of the current tendency to classify as a P Cygni star any star whose spectrum shows at least one absorption line with a red emission component. Thus some authors will even classify α Cyg, a typical stable supergiant, as a P Cyg star, since its H_α line shows a weak emission component in high-dispersion spectra. It should be remembered that the cB1eq spectrum of P Cyg itself shows extremely bright and wide lines of hydrogen, helium, and other elements, the absorption components showing shifts of from −28 to 300 km/sec. The spectrum of P Cyg (Figure 116) differs from the spectra of typical novae in the fairly early stages of the declining branch only in that the lines are somewhat narrower.

About a dozen stars with P Cyg spectra were discovered in the Large Magellanic Cloud, where they are among the brightest stars. Their absolute magnitude is on the average close to -7^m on the new scale of distances. They include the brightest known star S Dor, with absolute magnitude of $-8^m.9$ at the maximum. This is an eclipsing variable, however. Both its components are apparently P Cyg stars, with absolute magnitudes of close to $-8^m.2$. They are both small-amplitude irregular variables. There are certain indications that the absolute magnitude of P Cyg also ranges from -5^m to -7^m. The absolute magnitudes of such individual stars naturally cannot be determined with any reliability in our Galaxy. A number of stars exist, however, which possess a P Cyg spectrum and variable magnitude, although they are not supergiants. At any rate many of these stars have apparent magnitudes of $10^m—11^m$ and are located at fairly high galactic latitudes, so that their luminosity should be fairly low.

The very weak nova-like companion of R Aqr occasionally shows a P Cyg spectrum; the companion of o Cet, a white or intermediate dwarf with absolute magnitude of $+6^m$, and other stars also fall under this category.

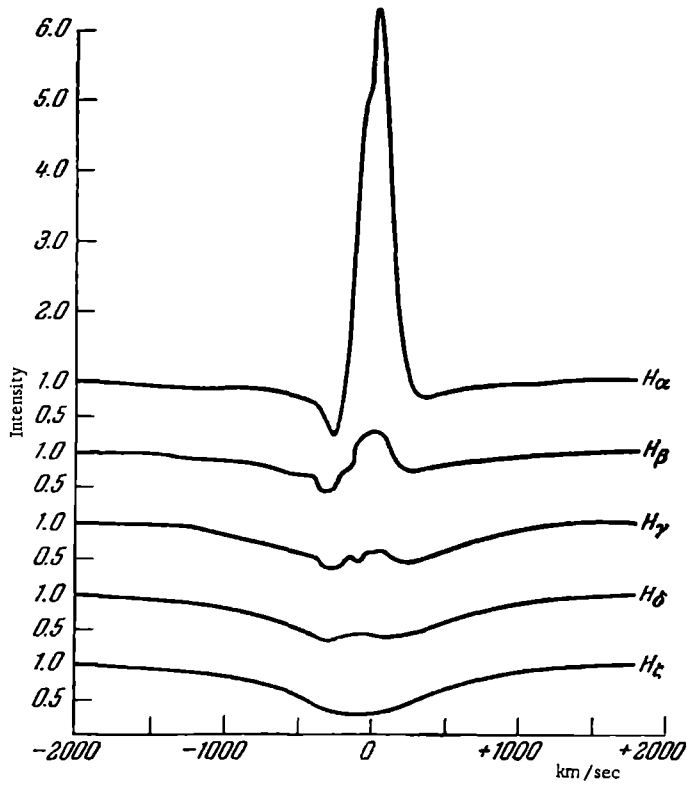


FIGURE 116

Most of the P Cyg stars are of B or O type. Beals (1951) counted 69 known P Cyg stars in the Galaxy and the Magellanic Clouds; his list includes Deneb (cA2keaq), Z And, the companions of R Aqr and o Cet (B8eq), ε Aur (F2eaq), Rigel (B7eaq), η Car, μ Sgr (B8eβq), β Lyr and others.

This generalized population now includes a great variety of stars, whereas the original P Cyg stars were B-type supergiants with numerous bright and fairly wide lines with dark components. If we stick to the original definition of P Cyg stars, the Large Magellanic Cloud, apart from S Dor, contains but a single variable of this group. The other 8 stars observed in this galaxy are not variable, like numerous galactic stars of this type.

No one would have ever thought of classifying the P Cyg stars as novae had it not been for the curious history of P Cyg itself. In 1600 this was a 3^m star, and earlier it apparently had been even fainter. By 1625 it had dropped to 6^m and then progressively grew fainter. In 1655 its brightness increased to $3^m.5$ and after that it again became "invisible." Thus fluctuating between "invisibility" and 6^m, it steadied at $4^m.9$ in 1715, and minute brightness fluctuations can now be detected only with an electrophotometer.

Other stars with similar spectra probably also underwent similar changes, although this is not quite plausible for such stars as Deneb, Rigel, ε Aur, and β Lyr.

P Cyg stars are apparently supergiants with highly extended and unstable atmospheres in a state of steady and fairly slow expansion. Occasionally the ejection of surface matter sustaining the extended atmosphere subsides and the underlying normal photosphere supplies a substantially reduced luminous flux: the star thus grows fainter as its extended atmosphere is temporarily dissipated. If this is indeed so, the phenomena occurring in the atmospheres of P Cyg stars seem to be the exact opposite of the effects in the atmospheres of novae. The nova in its normal state is characterized by total absence of any dense extended atmosphere, whereas P Cyg stars are normally shrouded in an extended atmosphere which is occasionally dissipated. P Cyg stars should be closer to Be stars, or to B stars in general, rather than to novae. This conclusion is also supported by the frequent variations in the magnitude and the spectrum of Be stars and the existence of a whole range of stars with transitional spectra between P Cyg and Be types. For example some stars, variables included, show Be spectra with bright Fe II lines. One of these stars, XX Oph, often shows very bright lines of the same structure as P Cyg; its declining phase takes substantially less time than the phase of constant maximum luminosity. Similar effects are observed for Z CMa. In the spectrum of RW Peg (a B-type star) all the Balmer lines are very bright; in other stars, only some of the Balmer lines are bright, mostly H_α and H_β or only H_α, which is sometimes observed only as a weak component and occasionally disappears.

We are thus led to the following general conclusion: B-type stars are often characterized by transient or steady conditions which favor the formation of fairly extended atmospheres of variable height, density, and stability. When these atmospheres are relatively tenuous and almost stationary, they fluoresce in accordance with Rosseland's theorem: hydrogen atoms undergoing recombinations emit in a number of Balmer lines. These atmospheres, "suspended" like the reversing layer of the Sun, are probably made up of pure hydrogen.

P Cyg is known by its anomalous yellow color and energy distribution in the continuum which corresponds to a much lower temperature than its spectrum. According to some authors, the temperature of P Cyg inferred from the energy distribution in the continuous spectrum was found to fall from 8300 to 5700° on passing to longer wavelengths, whereas its spectral type corresponds to temperatures near 20,000°. This effect is in good qualitative and quantitative agreement with Kozyrev's theory of extended atmospheres, which leads to similar anomalies in energy radiation. Recent findings seem to indicate, however, that the anomalous yellow color of P Cyg is entirely attributable to interstellar absorption. Other stars of this type have not been studied, and the question of the true color of P Cyg remains open.

The theory of line profiles in the spectra of P Cyg stars has been developed in some detail. It considers absorption, radiation, and scattering processes in an expanding atmosphere, and the theoretical results show a good fit with observations.

The mass lost by a typical P Cyg star following the expansion of its atmosphere is estimated at about 10^{-5} solar masses annually. This figure is determined by the rate of expansion, the size of the star, and the density of the atmosphere (a quantity which is not easily determined). The high rate of mass loss indicates that if P Cyg stars have been ejecting stellar matter continuously, they could not exist for more than 10^5—10^6 years. No direct mass determinations are available for these stars. The mass of the eclipsing binary S Dor is found to be between 60 and 160 solar masses from its luminosity and general statistical relations for spectroscopic binaries.

Note, however, that most of the stars grouped under this category, e.g., some stars in Perseus, are characterized by an extremely low rate of ejection, as is evident from the presence of a single bright hydrogen line in their spectra. Their mass loss rate is therefore thousands of times less than for typical P Cyg stars. If the mass ejection rate is markedly decreased along the declining branch of the star (which took nearly a century for P Cyg), most of the stars classified under this category may have easily existed for 10^9 years and longer in this phase. They are thus older than many of the ordinary stars, provided that their age is not limited by considerations connected with the release of radiant energy.

§69. **Late-type nova-like stars.** Late-type variables, frequently associated with nebulae, and in particular RW Aur and T Tau stars, are sometimes classified as nova-like stars. All these stars are of spectral types F5—G5 and are apparently subgiants. They are regarded as nova-like stars because R Mon and T Tau show low-excitation bright lines (H, Ti, and Fe II) with violet absorption components. In the spectrum of T Tau these components show a shift of −165 km/sec. It is not clear whether this shift is associated with ejection of matter from the atmosphere or simply the presence of ascending atmospheric currents, since the observed velocity is much less than the parabolic velocity at the surface. These stars are surrounded by small weak nebulae and are physically connected with variable comet-like nebulae. Other stars with similar variations have pure absorption spectra and are not associated with any bright nebulae. If the spectra of these stars show emission lines, these lines grow stronger toward the maximum, at variance with the behavior of emission bands in the spectra of novae. These stars, grouped under the same category only according to photometric signs of irregular light variation, are apparently physically different, and the main difficulties are associated with the variability of their spectra which is linked with light variation and has been hardly studied. T Tau stars have recently attracted much attention: these are extremely young stars forming associations (as discovered by Ambartsumyan).

At any rate it is clear that the presence of bright spectral lines is definitely insufficiently for classifying a star as nova-like. Stars with bright lines may be of fundamentally different nature, with different atmospheric processes and different origin of the bright lines, and we still do not know what general classification they fit.

For example, it is significant that stars of the above type are not markedly different in their spectra from other stars described below. Moreover, UZ Tau and UV Cet stars are similar to the rest in another respect, namely the irregular light fluctuations.

UZ Tau is a dwarf, with its spectrum varying from dM3 to continuous. As the star approaches the maximum, the spectrum develops numerous bright lines, some of them forbidden. Similar spectra with bright lines of various intensities and multiplicities, though without nebular lines, were observed for other dMe dwarfs related to UZ Tau; these were YY Gem (the companion in the double system of Castor), HD 196982, BD + 19°5116, and others.

The "flaring" UV Ceti variables, undergoing occasional brief flares, also show similar spectra. Proxima Centauri (a component of the α Cen system) is close or actually related to these stars. There is an as yet imperceptible transition from this group to dMe stars of constant brightness, which sometimes show only bright components near the calcium H and K lines. T Tau stars and the weak variables in the Orion Nebula, showing spectra with a wide range of emission intensities, are apparently all related to these stars also. Some of the weak stars in the Orion Nebula have spectra with periodically appearing and disappearing strong H_α emission line, although the brightness of the stars remains constant. While some of the late-type dwarfs with dMe spectra are situated in bright or dark nebulae, others (Proxima Centauri) are not associated with any nebulae.

Besides late-type dwarfs and subgiants with emission lines and "nova-like" features in their spectra, there are also G—M giants and supergiants with bright, sometimes forbidden, lines; these include VV Cep, ρ Cas, WY Cem, Ross 1985, W Cep, WY Vel, W Ser, and others. In some cases these stars probably have an invisible or almost invisible hot and nova-like companion, which is responsible for the bright lines and their variation. This may not be always the case, however. At any rate, the forbidden bright lines which often require high excitation and ionization, are observed not only in red giants but also in red dwarfs, where a hypothetical low-luminosity hot companion would immediately reveal itself by a continuous spectrum superimposed on the spectrum of the cold M-type dwarf.

There is a possibility that the recently discovered nebulous objects of extremely low luminosity are in fact the hot and invisible sources exciting the emission in nebular and other bright lines in the spectra of systems comprising a cold dwarfs and a giant. Thus, near NGC 1999 Herbig and Haro discovered 7 nebulous objects, the brightest with stellar magnitude of $17^m.5$, and the rest much weaker. It is not clear at this stage whether these nebulous spots a few seconds of arc in diameter contain an even weaker star or whether they are not stars at all in the usual sense of the word.

At any rate, if these are stars, they are neither red nor yellow. They, or more properly the surrounding nebula, give an extremely weak continuous spectrum on which are superimposed a very bright H_α line, weaker [O I] and [S II] lines, and still weaker [O II], Ca II, [Fe II], [Ne III], and other lines. This spectrum is similar to that of the T Tau nebula. The absolute stellar magnitude of these mysterious objects is below 9^m (assuming that they are at the same distance as the apparently nearby Orion Nebula), and the absolute magnitude of the source of continuous emission in these objects is at most $+12^m$ (assuming that the continuous spectrum is not emitted by the smallest of these nebulae).

In summing up we see that there is a continuous gradation of stellar spectra which mostly belong to irregular variables and to a certain

(possibly very small) extent are reminiscent of the spectra of novae. These spectra merge either with the spectra of normal hot supergiants of types cA and cB, or the spectra of hot dwarfs and ultradwarfs, or else the spectra of medium-type subgiants and spectra of red giants and dwarfs.

The common feature of these stars with novae is the presence of extended atmospheres with ascending currents of hot gases, which as a rule do not produce ejection of stellar matter. They all should have a source of more or less powerful ultraviolet radiation. A hot star generally (though possibly not always) acts as a source of this radiation. In some nova-like stars it is hot and solitary, whereas in others it is only a weak and often invisible component of a system.

If novae are regarded only as very hot stars of medium and low luminosities which flare occasionally and eject an extended envelope in the eruption, nova-like stars should include only the Z And variables. Most of the other stars described as nova-like in the preceding are of fundamentally different physical nature. The only common feature for all the nova-like stars in the broad sense of the term is the presence of more or less extended or nonstable atmospheres or possibly nebulous envelopes. Further seeing that the spectra of cepheids and some eclipsing variables also develop or contain emission lines, we conclude that solitary and multiple stars of a great variety of types are characterized by an extraordinary diversified selection of atmospheres which are more extended and more active than the Sun's atmosphere. The evolution of large classes of stars is apparently governed by additional factors which are still unknown.

§ 70. **The mechanism of nova eruption.** The theory of novae is often regarded as the theory of various processes taking place during the eruption, and not the theory of the eruption as such. Some of these theories have been developed in considerable detail, e.g., the theory of Ambartsumyan and Gordeladze which explains the light curve of novae as the result of expansion of an optically thick envelope whose optical thickness falls abruptly near the maximum and the star starts showing through the gas. An adequate theoretical model has been developed for the abrupt drop in brightness followed by a slow recovery as observed for N Her 1934, and a number of other effects.

As regards the eruption proper, a number of hypotheses have been advanced, but it is not clear whether the eruption is confined only to the outermost layers of the star or affects the interior also. At any rate, it has been established that the eruption is caused by quasiperiodic buildup of instability in stars of a certain type: these are extremely hot stars with a wide range of luminosities, which invariably fall between the luminosities of white giants and white dwarfs.

According to some hypotheses, this instability is associated with external factors: planets impacting on the star surface, interaction with nebulae, etc. Other hypotheses link the development with interior processes—increase of temperature following nuclear reactions, and so on. Lebedinskii developed the mathematical theory of a shock wave propagating in the stellar interior, which strips a symmetrical or asymmetrical envelope off the star on reaching the surface.

According to Birman, the instability develops at a certain depth where radiant equilibrium changes over to adiabatic equilibrium. In this case

the star should be poor in hydrogen and helium, and the occurring changes lead to a release of ionization energy.

Another theory, also developed in some detail, is that of Gurevich and Lebedinskii, who reach the conclusion that a medium-luminosity hot star will readily erupt on reaching an interstellar hydrogen cloud with an increased deuterium content. The deuterium accreting to the star produces a zone of nuclear reactions which lead to energy excess followed by thermal explosion at a certain depth below the surface. The various theories, however, ignore the existence of the white-blue sequence on the HR diagram, where the novae for some reason join on one end the hot giants and on the other the white dwarfs. Vorontsov-Vel'yaminov has suggested that this sequence, or at least part of it, reflects the true course of evolution.

If we accept the inevitability of a decrease in luminosity with decreasing mass, the luminosity of novae must diminish as they lose mass in successive eruptions and by continuous ejection of matter between the eruptions, in the Wolf—Rayet stage. It therefore seems highly probable that an erupting nova moves down along the white-blue sequence, and its eruptions become less frequent in virtue of the amplitude—frequency relation. Novae with particularly large amplitudes which erupt very seldom, at intervals of millions of years, are already blue dwarfs. The gradual evolution of novae into stable white dwarfs can hardly be questioned.

Less certain is the starting premise of Vorontsov-Vel'yaminov, who maintains that some of the hot giants develop into recurrent novae on passing through the Wolf—Rayet stage. It is not clear what giants are capable of this development. Presumably, suitable candidates are those with the highest masses or possibly hydrogen-poor giants. Note, however, that the galactic concentration of Wolf—Rayet stars is much more prominent than that of novae; moreover, the Wolf—Rayet stars are mostly binaries. This is possibly a side line in the evolution of hot and massive supergiants, which have thus found a way to get rid of the excess mass. Yet some of the absorption O-type stars are known to be poor in hydrogen, e.g., HD 160641. This star has an anomalous radial velocity, +100 km/sec, and the members of this group are therefore expected to show lower galactic concentration than the ordinary O-type stars (approaching the concentration of novae). Numerous hot stars have been discovered in high galactic latitudes; they are possibly related to the low-luminosity stars occasionally found in globular clusters where novae erupt from time to time.

The evolution of novae deserves special attention, since so far this is the only class of stars which evolve while losing mass at a high rate that can be measured directly, whereas the substantial corpuscular loss of mass by other stars has never been observed in fact.

§ 71. **Supernovae.** Stars which flare suddenly to a very high maximum (much higher than the maximum of typical novae) and then decline gradually like novae are known under the unfortunate misnomer of supernovae.

No galactic supernova has been observed after Tycho Brahe, who witnessed a supernova explosion in 1572. In retrospect, we can assign to galactic supernovae the stars which erupted in 1604 in Ophiuchus, in 1572 in Cassiopeia, in 1054 in Taurus, and some other stars, which like the supernova in Taurus are described only in ancient chronicles. These reports naturally do not give reliable information on the luminosities of supernovae or on the dispersion of their luminosities. Telescopic

astronomy, however, has recorded a few dozen supernovae in other galaxies. According to these observations, the average absolute magnitude of supernovae on the old distance scale is -14^m, the individual magnitudes fluctuating from -10^m to -18^m. It is not clear to what extent this dispersion of luminosities is significant and whether or not the supernovae constitute a direct continuation of the ordinary novae, whose upper-bound magnitudes reach -10^m and even -11^m. This difficulty stems from the fact that we can never be certain of catching the exact maximum of a supernova; moreover, the distances to other galaxies are known with very low accuracy. We recall that even the distance modulus of the Andromeda Nebula has a probable error of $1^m.5$. The available dispersion estimates of the luminosities of galactic novae are also uncertain.

Yet the light curves of novae are seldom reminiscent of the gradual and slow curves of supernovae, and the complete dissimilarity in their spectra points to fundamental qualitiative differences between novae and supernovae, despite the fact that the luminosity of some supernovae exceeds the luminosity of the brightest novae at the maximum by less than two orders of magnitude. Zwicky (1962) identified five different types of supernovae.

The light variation amplitude of supernovae and their luminosities before and after the eruption are not known, since individual stars are never visible in the other galaxies at these phases, and in our Galaxy the pretelescopic supernovae have not been identified with any certainty.

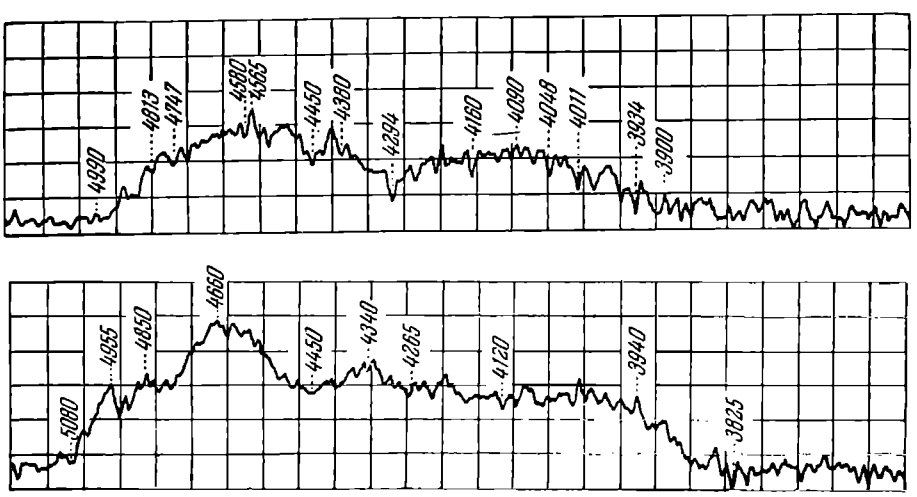


FIGURE 117

We distinguish between two types of spectral variations in supernovae; type I has been studied in greater detail than type II. The studied range covers amplitudes of up to 8 mag. Near the maximum the continuous spectrum of type I supernovae shows no distinct lines, and no absorption lines are observed altoghether. Later, very wide emission bands appear, which cannot be identified with known spectral lines. Figure 117 shows the tracing of the spectrum of a type I supernova in NGC 1003 51 days and

339 days after the maximum. The band widths show the emission originates in gas expanding with velocities of about 6000 km/sec.

The ultraviolet end of the spectrum is weak, and it grows weaker with time. In the visual spectrum the unidentified emission bands change their intensity and structure. Six months after the maximum, red lines 40 Å wide appear, which are identified with [OI]. Their intensity increases, and eventually they remain the only observable feature in the spectrum.

In the photographic region, the structure and the brightness of bands do not change, but the band centers invariably shift toward the red, the shift reaching 150 Å after one year.

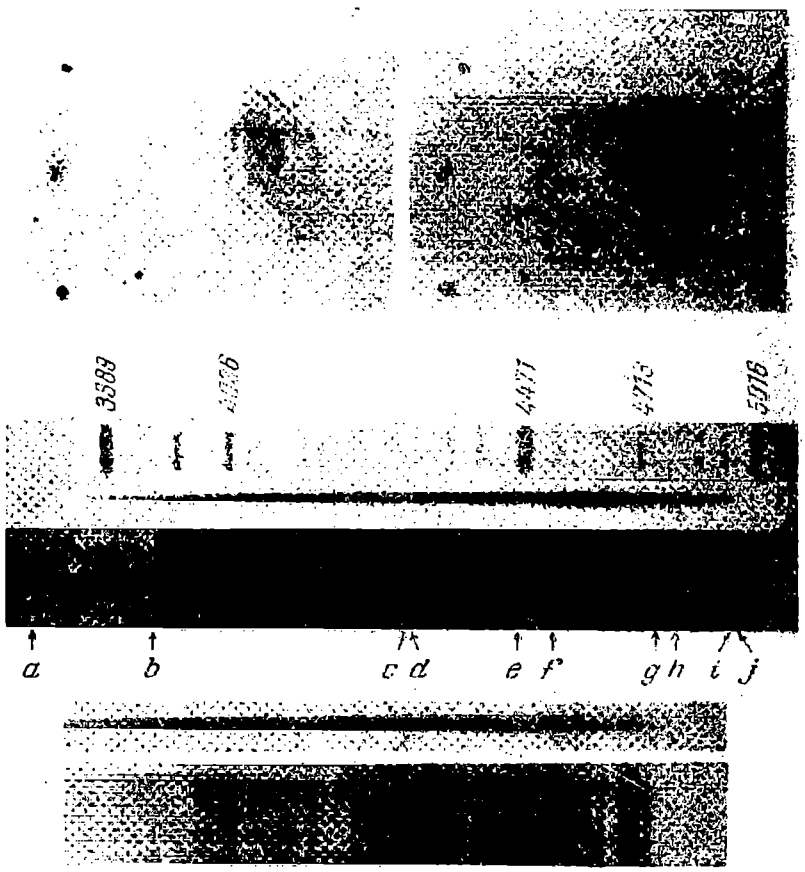


FIGURE 118

The attempts to explain the type I spectra as a superposition of the parabolic band profiles of typical novae, assuming broadening of up to ± 6000 km/sec, are not very convincing. The spectra of type II supernovae have a bright ultraviolet end, and the absorption and emission lines in the spectrum are more or less identifiable with the lines of hydrogen, ionized

nitrogen, and other elements appearing in the spectra of typical novae. (Figure 118 shows the original and the broadened spectrograms of the supernova of 1936 in NGC 4273 and of the supernova of 1926 in NGC 4303. The photograph on top is NGC 4273; the right photograph shows the supernova.) The spectra of type II supernovae are nevertheless definitely different from the spectra of typical novae. The luminosity of type II supernova is on the average two stellar magnitudes less than the luminosity of type I supernovae.

The paucity of data leads to contradictory conclusions concerning the frequency of supernova explosions. The most probable estimates are apparently the following. Type I supernovae are related to Baade's Population II stars, and type II supernovae mostly explode in Population I regions. Type II supernovae (9 in number) were observed only in Sb and Sc galaxies, whereas type I supernovae occur in galaxies of all kinds, their frequency, however, increasing in Sc and E galaxies. In an average galaxy, a type I supernova explodes on the average once every 2000 years, and in a giant galaxy, like the Milky Way, the rate of occurrence is once every 400 years. Type II supernovae are 6 times more frequent, exploding once every 30 years in a giant galaxy. There are naturally enormous fluctuations in these figures. Thus, in our Galaxy three type I supernovae exploded in 500 years in 1% of the galactic volume, and not a single type II supernova was observed. On the other hand, at least two supernovae were observed in three other galaxies within a few decades.

The fate of type II supernovae is completely unknown, but type I supernovae often produce peculiar rapidly expanding nebulae. The size of these nebulae, however, often varies between wide limits.

The spectrum of the Crab Nebula in Taurus is entirely different from the typical spectra of planetary and diffuse nebulae. More than 80% of the total energy is radiated in the continuous spectrum. The weak hydrogen lines indicate that the nebula is poor in hydrogen. The emission lines in the central part of the nebula are split, revealing a radial expansion with velocities of about 1000 km/sec. Figure 119 is the spectrum of the Crab Nebula showing the split lines. Comparison of direct photographs taken after a prolonged interval reveals definite expansion of the nebula. Assuming uniform expansion, we find that it began some 900 years ago (in other words, this is the age of the nebula). The position of the Crab Nebula in the sky is very close to the reported position of the supernova of 1054, described in Chinese and other chronicles.

It is thus fairly certain that a supernova exploded in Taurus in 1054 and its position is now marked by an unusual expanding nebula with an inconspicuous white central star of 16^m, which is probably the residue of the star that exploded nine centuries ago.

Baade's photographs taken through narrow-band filters show that the continuous spectrum is emitted by an extended luminous mass of amorphous shape (Figure 120). The gaseous spectrum, on the other hand, is emitted by thin filaments which cut through the amorphous mass in the tangent plane and form a sort of a network enclosing a sphere (Figure 121).

The nebula, and not the central star, is one of the most powerful radio sources in the Galaxy. The continuous spectrum emission and the radio waves of the nebula constitute the bremsstrahlung of relativistic

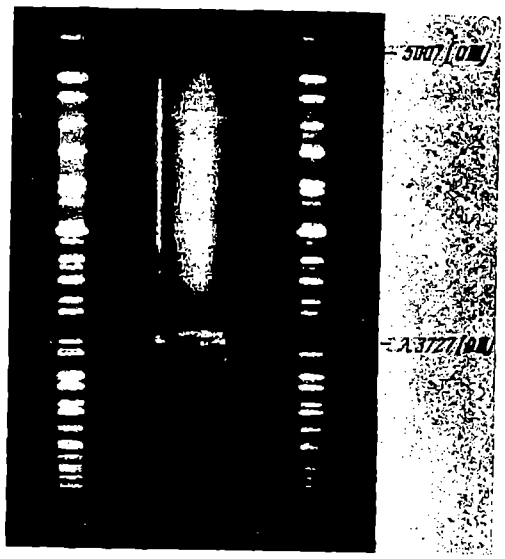


FIGURE 119

FIGURE 120

(exceedingly fast) electrons in the magnetic field of the highly ionized nebula. This hypothesis was confirmed by Dombrovskii and Vashakidze, who observed the theoretically predicted strong polarization of light from this nebula. The central star is not necessarily identified with the ex-supernova. It has a weak yellow neighbor. The temperature of the ex-supernova cannot be less than 100,000°. Different mass estimates are available for the nebula, depending on the particular emission mechanism used. Its probable mass is apparently near 0.1 $M_\odot$.

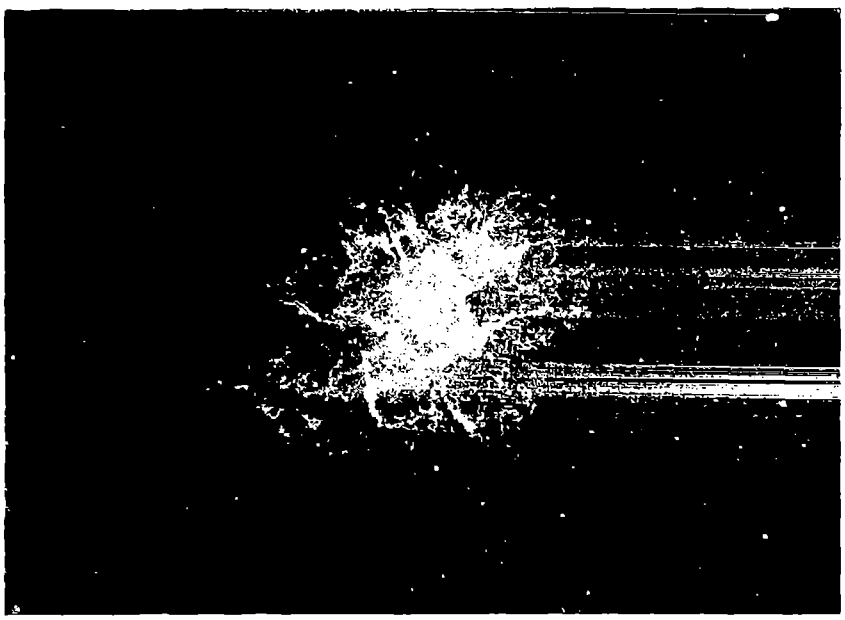


FIGURE 121

The peculiar nature of the Crab Nebula encouraged the search for other similar nebulae in regions marked by supernova explosions. Minkowski found a weak nebula near the site of the supernova of 1572. A radio source was also discovered there, although it was 10 times less powerful than the Crab Nebula. The site of the supernova of 1604 is also marked by a very small and very faint nebula, similar to the Crab Nebula in certain respects. Its spectral lines are virtually the same, although not split, but the continuous spectrum is weak. A conspicuous radio source was also identified with this nebula. The radio power, the brightness of the nebula, and the time after the eruption of the parent supernova are thus in no way related with one another, and the supernova explosions show a great variety.

The extraordinarily powerful radio nebula in Cassiopeia (Cassiopeia A) in Shklovskii's opinion is the remnant of the supernova of 369 described in ancient chronicles. The optical nebula 2'.5 in diameter identified with the radio source definitely expands and its random velocities reach about

1500 km/sec. This is consistent with the hypothesis of formation of unusual expanding radio nebulae on the site of supernovae, although no two such nebulae are alike.

The various hypotheses and theories of supernova explosions (e.g., the evolution of normal stars into collapsed neutron stars) are still largely unfounded because of the paucity of observational data.

It should be noted, however, that of all the known natural phenomena involving tremendous changes in a relatively short time (on the cosmic scale, that is), supernova explosions are the most prodigious and the farthest from anything that we can ever hope to accomplish in our laboratories.

REFERENCES TO CHAPTER XI

1. PAYNE-GAPOSCHKIN, C. The Galactic Novae.— Amsterdam. 1957.
2. CECCHINI, G. and L. GRATTON. Le stelle nuove.— Milano. 1941.
3. VORONTSOV-VEL'YAMINOV, B.A. Gazovye tumannosti i novye zvezdy (Gaseous Nebulae and Novae).— AN SSSR. 1948.
4. Observation des novae. Théories des novae. Supernovae. Paris. 1941.
5. AMBARTSUMYAN, V.A. et al. Teoreticheskaya astrofizika (Theoretical Astrophysics), Ch. V.— Gostekhizdat. 1952.
6. MUSTEL', E.R. — Various papers in Astron. Zh., 1944—1947.

Part Four
DIFFUSE MATTER

The upper-bound value obtained from dynamic considerations for the overall mass of diffuse matter in the Galaxy is equal to the combined mass of the stars. The actual mass of the diffuse matter, however, is one or two orders of magnitudes less than the combined stellar mass (radio observations show it to be 2% of the total galactic mass), but even so it is a substantial component of the Galaxy.

Diffuse matter occurs mainly in two forms: fine dispersed dust and low-density gaseous matter. Both forms of diffuse matter exist as discrete formations — nebulae — and as a continuous background of highly variable density. The two extreme schools of thought regard the interstellar medium (1) as a perfectly continuous medium, and (2) as being completely composed of numerous discrete, but small, nebulae or clouds. In fact the bulk of the diffuse matter is apparently distributed as a continuous but highly inhomogeneous substrate with various motions and condensations of a variety of scales and densities. However tenuous the diffuse matter in certain parts of the Galaxy, it imperceptibly merges into even more tenuous diffuse matter which fills the intergalactic space.

All forms of diffuse matter show a pronounced galactic concentration.

Galactic nebulae are divided into gaseous and dust nebulae in terms of their physical nature and into bright and dark nebulae in terms of their external appearance. Dust nebulae are either bright or dark, but they all have an irregular, diffuse outline. Nevertheless, certain dark nebulae, known as globules, have very small diameters and fairly sharp, though not always circular, outline. Among the bright gaseous nebulae some are diffuse whereas others are very small and sharply outlined: these are the planetary nebulae with a weak star — the nucleus of the nebula — at the center. Planetary nebulae show a lower galactic concentration than the rest of the diffuse matter, and like the dust globules they constitute the most discrete formations of the diffuse matter.

The interstellar gas is also observed using its characteristic absorption lines in the spectra of distant hot stars and a variety of radio-astronomical techniques.

B. A. VORONTSOV-VEL'YAMINOV

Chapter XII

GASEOUS NEBULAE

§ 72. Description of planetary nebulae. There are over 600 known planetary nebulae at present, mostly discovered by Minkowski in the 1940's. They show a moderate concentration toward the galactic plane, like the novae (Figure 122). They are greatly concentrated toward the center of the Galaxy, again like novae, and are probably part of the constitution of the galactic nucleus. In globular clusters there is one known nova and one known planetary nebula.

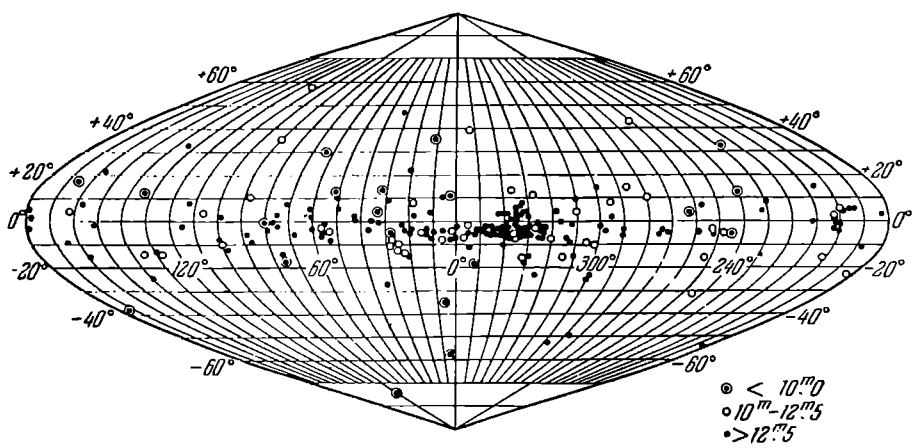


FIGURE 122

The integrated photographic magnitudes of the known planetary nebulae range from $6^m.5$ (NGC 7293 in Aquarius) to 18^m and higher. Visually they are on the average 1.5 mag brighter, since they are greenish in color, and occasionally reddish due to interstellar absorption. The apparent diameters of the planetary nebulae do not exceed 15' (the Snail in Aquarius). The diameters of numerous nebulae are 5—15" and less, so that a considerable number of planetary nebulae do not show any visible disk. These are star-like nebulae, which can be distinguished from stars only spectroscopically: the characteristic gaseous spectrum generally contains the forbidden green [OIII] lines.

The average surface brightness of planetary nebulae ranges between limits. It is correlated neither with the integrated brightness nor with t angular diameter of the nebulae.

The planetary nebulae are small, so that their structure can be resol only with powerful instruments.

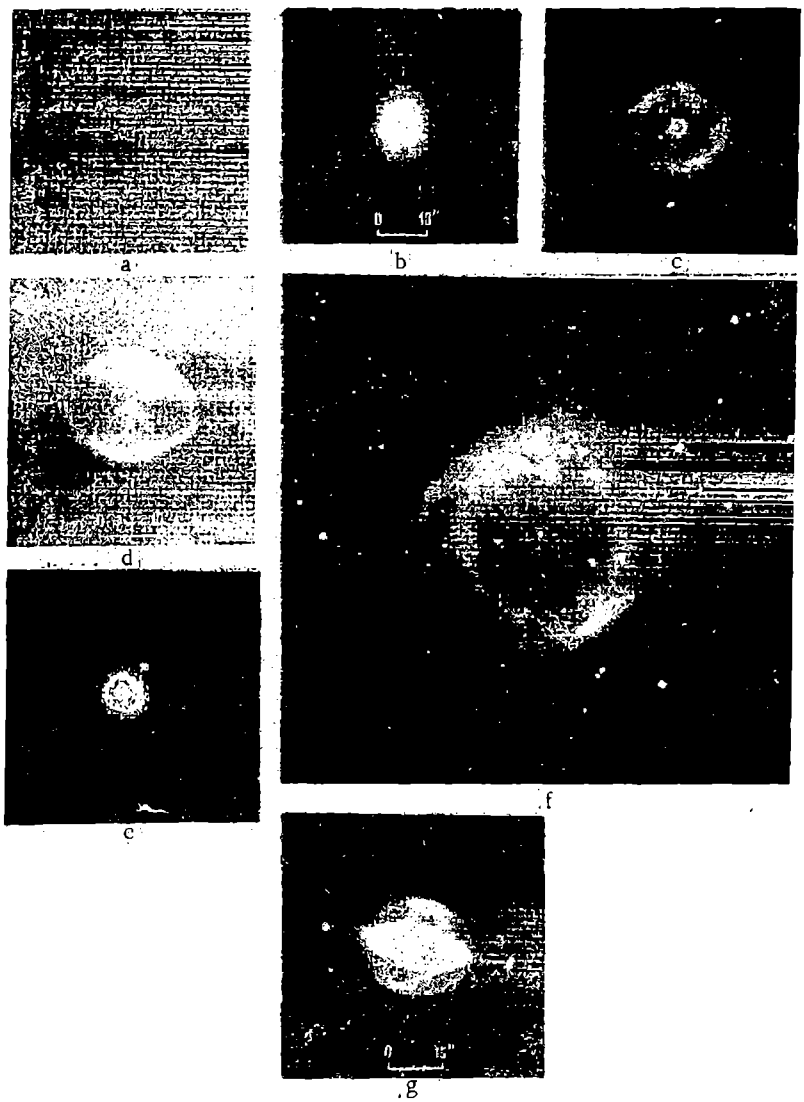


FIGURE 123

Morphologically the planetary nebulae are classified into the following groups (Figure 123):

I — star-like (a),
II — regular illuminated disk,
IIa — with brightness concentration toward the center (b, NGC 6572),
IIb — uniform brightness,
III — disks with nonuniform brightness,
IIIa — patchy structure (c, NGC 1501)
IIIb — traces of a ring (d, NGC 6058)
IV — ring nebulae (e, NGC 7293, f, NGC 1535),
V — transitional type between planetary and diffuse,
VI — anomalous nebulae (g, NGC 7009).

The unusual shapes under Category VI are quite significant. Some nebulae are composed of two envelopes — a bright inner envelope and a weaker outer envelope. The bright ring is often seen against the background of a weaker disk of fairly moderate brightness. Ring nebulae are somewhat elliptical in shape. The Snail Nebula appears as two intersecting rings of equal diameter.

A nucleus star is situated at the center of each planetary nebula. This is generally a weak star, which is often very difficult to detect. The nucleus is often weaker than the surrounding gas envelope.

The nuclei of planetary nebulae are the hottest known stars. Their spectra are absorption O type, Wolf—Rayet type (with bright emission bands), or possibly continuous, without any lines. The temperatures of the nuclei also increase in this order, but they are never less than 30,000°. The central position of the nucleus leaves no doubt as to the genetic relation between the star and the nebula.

§ 73. **The spectra and the light of planetary nebulae.** Slit spectrographs reveal numerous spectral lines and make it possible to determine the radial velocities of the nebulae. Slitless spectrograms give the distribution of various wavelengths inside the nebula.

The nebular spectra are very diversified (Figure 124), which points to different degrees of ionization and excitation in the nebular envelopes. Some spectra show a predominance of hydrogen lines, as in diffuse nebulae. The most typical feature of the planetary nebulae, however, are the strong green lines, originally identified with a hypothetical element "nebulium," which are now known to be the forbidden lines of doubly ionized oxygen [OIII]. All in all, the number of forbidden lines in the spectrum is often greater than the number of observed allowed lines. The continuous spectrum sometimes does not show or is extremely weak; often it is observed only beyond the Balmer limit, i.e., it is produced by recombination of protons with electrons followed by transition of the captured electron to the second level. The nebular spectra as a rule contain lines of one chemical element in different stages of ionization: e.g., [OI], [OII], and [OIII]. This is never observed in stellar spectra under conditions of thermodynamic equilibrium.

The nebular spectra show the presence of H, He, C, N, O, S, Ne, F, Cl, A, Fe. The relative intensities of the lines of these elements vary between wide limits.

The distribution of different wavelengths in nebulae is distinctly stratified. On the whole, the higher the ionization potential of the atom or the ion the smaller is the radius of the corresponding emission zone

around the nucleus. Stratification considerations, however, should allow for the actual structure of the nebula, which so far has been ignored. The forbidden lines of atoms with high ionization potentials and low excitation potentials, on the other hand, are emitted approximately in the same regions with the hydrogen lines, since they are excited by collision with free electrons, supplied mainly by the almost completely ionized hydrogen.

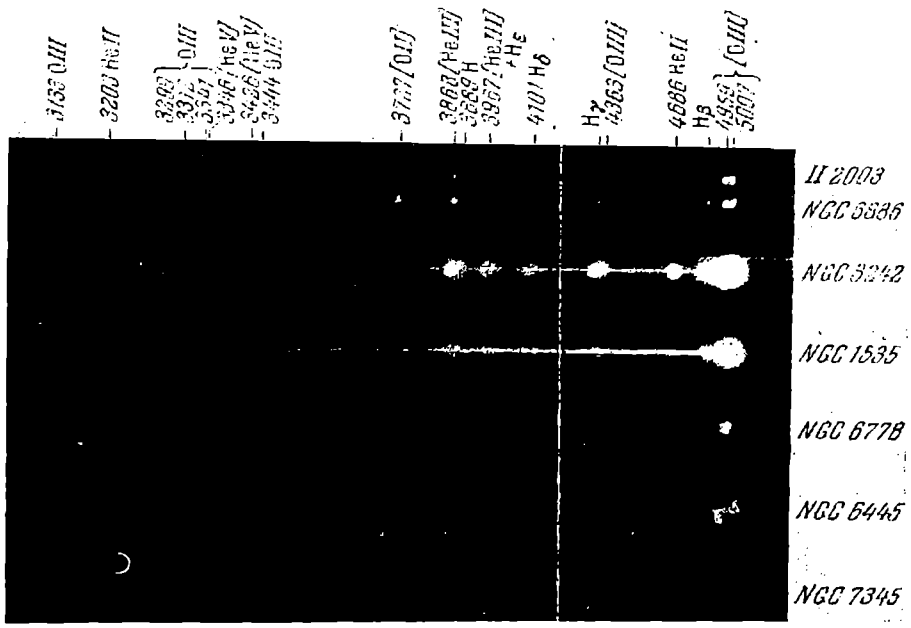


FIGURE 124

The nebular lines are narrow, so that the small-scale turbulence in the nebulae is insignificant, although the electron temperature reaches about 10,000°. Large-scale turbulence sometimes reaches velocities of a few kilometers per second.

The spectral lines observed with the spectrograph slit aligned along the diameter of the nebula often show a definite splitting, which increases toward the apparent center of the nebula (Figure 125). Zanstra's is the only correct interpretation of this effect: he attributes it to radial expansion of the nebulae with velocities from a few km/sec to 40 km/sec. It exceeds the parabolic velocity at the corresponding distance from the nucleus, and the nebula will continue expanding at this rate until it is completely dissipated, unless the drag of the medium or complex radiation pressure effects decelerate it sufficiently. Both the drag and the radiation pressure are unknown factors. The effect of radiation pressure for real nebulae made up of a mixture of gases is highly complex because of stratification, certain differences in velocities, and possible velocity gradients in the nebula.

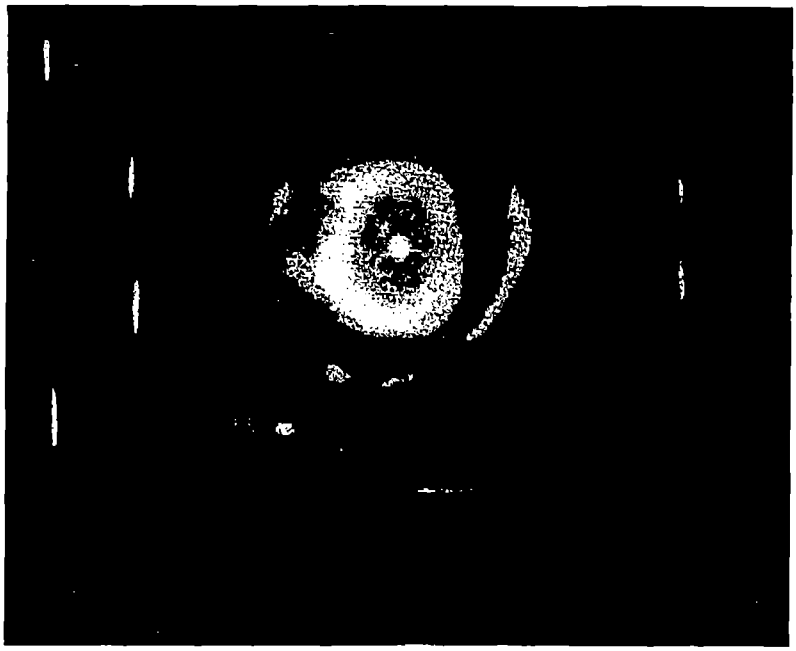


FIGURE 125

The light from the receding, rear hemisphere of an expanding nebula also reaches the observer since the nebular gas is perfectly transparent, especially in the forbidden lines. The spectral lines are therefore not only curved but actually split.

The age of planetary nebulae, assuming constant expansion velocity, if often as low as a few thousand years; direct measurements of the angular expansion velocity from photographs often confirm this conclusion.

The line intensities in nebular spectra, in particular the Balmer decrement, the intensity and the energy distribution in the Balmer continuum, are adequately explained by the theory of bright gaseous nebulae advanced in Chapter XVI. Analysis of spectroscopic data gives the mass of the envelope, its quantitative chemical composition, electron temperature, the temperature of the nucleus, the density of the envelope, and the nature of various processes in the nebula.

Note that the theory of diffuse gaseous nebula and extended envelopes around certain stars is largely based on the theory of planetary nebulae, mainly developed by Menzel, Aller, Zanstra, Ambartsumyan, and Sobolev. The reason for this is that various processes take place in planetary nebulae in their pure form, without interference from other factors; the exciting star is situated symmetrically at the center of the nebula, and the nebula itself is fairly symmetric in space. The spectra of planetary nebulae are also more accessible.

Without going into details, we will only indicate that the source of nebular luminosity is the hot nucleus at the center of the nebula. The envelope absorbs almost completely the ultraviolet radiation emitted in large quantities by the hot nucleus. Each ultraviolet quantum of the nucleus is re-emitted in one of the allowed lines of the subordinate spectral series of hydrogen and helium. The light is emitted through photoelectric ionization of atoms followed by recombination with free electrons which then cascade to the ground level.

The forbidden lines are produced by a different mechanism. Sufficiently fast free electrons collide with some atoms exciting them to a low-lying state: a low excitation energy is required since this state is close to the ground state. The nebula will emit in these lines if the gas density is low and the gas is far from the star (high dilution of incident radiation).

The theory shows that certain lines can be excited by special random mechanisms (Bowen's mechanism), but it is not clear whether these mechanisms actually exist. Hydrogen in nebulae is 10 times more abundant than helium, and thousands of times more abundance than C, N, and O. The atomic content of other elements is negligible.

The intensity of the [O III] 4363 Å line in the spectrum of NGC 4997 dropped by a factor of 2.6 over a period of 40 years. This is apparently associated with the decrease of electron temperature as the nucleus becomes cooler.

The envelope mass of a typical planetary nebula is estimated at $0.01-0.1$ solar masses, and the density is $10^{-21}-10^{-20}$ g/cm^3. The free path is thus tenths of an astronomical unit, and the time between successive collisions is a few minutes. These conditions have never been accomplished in terrestrial laboratories. Spectroscopic study of processes in planetary nebulae made a much greater contribution to atomic physics than the theoretical study of white dwarfs with their enormous densities, which are also unattainable in laboratories.

§ 74. Distances, sizes, luminosities, and spatial structure of planetary nebulae. The distances to individual nebulae were generally determined by modifications of the method originally proposed by Vorontsov-Vel'yaminov. He established the low dispersion of the absolute integrated magnitudes of planetary nebulae. The mean absolute integrated magnitude is determined statistically, e.g., from the dependence of the radial velocity on distance for objects taking part in galactic rotation. The fairly constant luminosity of the expanding nebulae indicated that these were optically thick objects, whose hydrogen absorbed completely the ionizing ultraviolet energy of the nucleus. This argument, however, ignores the strong emission in the forbidden lines and the role of the nuclear temperature. The luminosity of a nebula apparently somewhat depends on the temperature of its nucleus: δ, the difference in the stellar magnitudes of the nucleus and the entire nebula, increases with the increase in nuclear temperature. The apparent angular diameters of nebulae provided a poor indication of distances, since the nebulae are steadily expanding and their ages are all different.

According to Vorontsov-Vel'yaminov (1951) for planetary nebulae

$$M_p = 0^m.04 - 0.22\delta$$

and on the average $M_p = -0^m.5$.

The distances are obtained by correlating the absolute magnitude M_p with the apparent magnitude m_p and correcting for interstellar absorption.

Unfortunately, the interstellar absorption is very uncertain at the large distances where most of the known planetary nebulae are observed.

The spatial distribution of the known nebulae is shown in Figure 122. The nearest nebula is the Snail in Aquarius, NGC 7293, which is distant 290 pc; its dimensions are 210× 260 thous. a.u. The smallest nebulae with a distinguishable disk have diameters of 2,000—3,000 a.u., but some star-like nebulae are apparently much smaller. Among nebulae with clear disks, type-II nebulae are on the average the smallest ($13 \cdot 10^3$ a.u.), and nebulae of types III—V are on the average larger ($67 \cdot 10^3$ a.u.). Some typical planetary nebulae have diameters greater than 1 pc.

Shklovskii used the model of an expanding nebula which emits only in hydrogen lines and concluded that soon after the beginning of expansion the nebula should become optically thin: it will become partly transparent to ultraviolet radiation of the nucleus and its brightness will rapidly fall. He therefore suggests that the distances to optically thin nebulae can be measured from their angular diameters, assuming fairly constant masses and linear dimensions and correcting for surface brightness. At this stage, we cannot decide which of the methods is to be preferred. Moreover, we cannot distinguish with any certainty between optically thin and optically thick nebulae. Some astronomers regard ring and double-envelope nebulae as optically thin. The Ring Nebula in Lyra, in particular, is included in this category. Wurm, however, obtained for this nebula an excellent fit for the temperature of the nucleus using different lines in the spectrum. This is impossible unless the nebula is optically thick. There are also certain indications that the envelope masses of the nebulae are spread between wide limits.

Although no universally accepted method is available for the determination of distances to individual planetary nebulae, their average luminosities, masses, and densities are nevertheless true to orders of magnitude.

Since the nebulae are perfectly transparent, especially in the forbidden lines, their spatial structure can be investigated from the apparent brightness distribution over the disk. However, because of the stratification of radiation in planetary nebulae (see Figure 126, which shows the isophotes of the planetary nebulae NGC 6543 (top) and NGC 7009 (bottom) at different wavelengths), it is better to work with a slitless spectrograph in monochromatic light. Vorontsov-Vel'yaminov found that the spatial density in NGC 6572 (associated almost entirely with protons) increases from 10^{-21} g/cm^3 at the periphery to $2.5 \cdot 10^{-20}$ g/cm^3 at a distance of 0.1 of the radius from the nucleus. The radius of the nebula was taken equal to $5.5 \cdot 10^3$ a.u. The nebula expands at a rate of 2.5 km/sec, and its mass is probably steadily replenished by gaseous matter ejected from the central star. It is thus clear that this nebula, and also all others, formed from the stellar matter ejected by the nucleus. The nucleus of NGC 6572 is a Wolf—Rayet star, i.e., this is an object which steadily ejects gas at very high surface velocities. The gas ejected by the star is rapidly decelerated in the envelope by some unknown mechansim. The deceleration is possibly caused by electromagnetic forces, whose significance in planetary nebulae is quite substantial.

The Ring Nebula in Lyra has a nucleus with a pure continuous spectrum, and the space between the nucleus and the interior part of the spherical envelope is a relatively high vacuum. The density in the envelope increases away from the nucleus, and then starts decreasing, dropping almost to zero

at the apparent periphery. The outer weak envelope of almost uniform brightness points to the existence of gas around the bright ring envelope, where the overall density decreases in the equatorial plane. However, the optically emitting thickness of the nebula is constant along any radius.

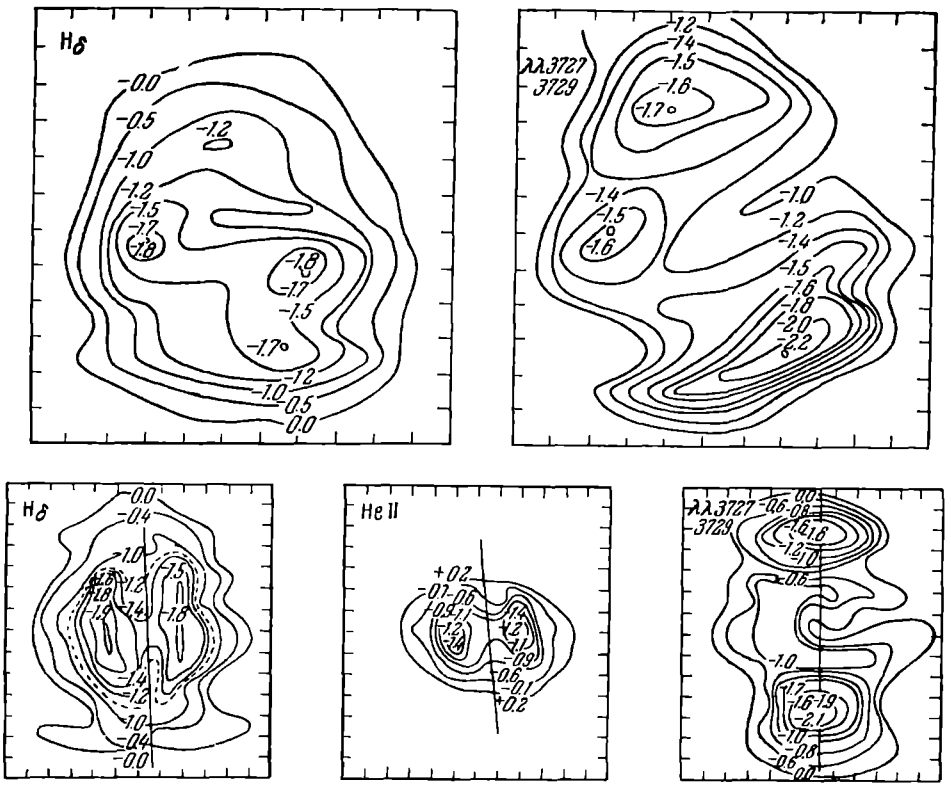


FIGURE 126

The photometry of nebulae with nonuniform distribution of brightness does not lead to unambiguous results concerning their spatial structure. Numerous nebulae have most peculiar shapes with various loop-shaped appendages and bright twin spots at the center, so that their actual spatial structure is a mystery. Curtis's study of the apparent shapes of nebulae, carried out at the beginning of the century, was extended by Minkowski in the 1950's, who studied the stratification of radiation in great detail. Gurzadyan attributes some peculiar features of brightness distribution in nebulae to the effect of strong magnetic fields which influence the motion and the density of gas. Magnetic phenomena are apparently the only factor which can explain the curious forms and loops observed in planetary nebulae.

§75. **The nature of the nuclei.** The temperature of nuclei is determined by Zanstra's method or one of its modifications (see Chapter XVI). The determination is based in the final analysis on the ratio of the luminosities

of the nebula in different spectral lines, which are the result of re-emission of the far ultraviolet energy of the nucleus. The temperature of the nucleus is thus inferred from a quantity not unlike the color index: we are comparing the energy in the visible spectrum with the far ultraviolet energy of the nucleus beyond the limit of the principal series of the particular atom. The temperature can be determined using the line intensities of different atoms and ions. The nucleus is assumed to radiate as a blackbody and the nebula is regarded as perfectly absorbing beyond the limit of the principal series of the particular atom.

The validity of these assumptions is not certain. For example, the energy in a given spectral region is absorbed by different species of atoms, and one atomic species cannot be regarded as absorbing the entire energy. The radiation of the nuclei may differ from blackbody radiation. The color index of a blackbody at infinitely high temperatures has the limiting value $-0^m.54$, whereas the color indices of some nuclei are lower. The color index of the nucleus of the Ring Nebula in Lyra is $-1^m.5$ or even $-2^m.0$. This is however, a highly complex problem, since the continuous spectrum of the nebula and its emission beyond the Balmer limit are often superimposed on the spectrum of the nucleus, lowering the color index. This is possibly the reason for the unusual color index of the Ring Nebula; spectrophotometric study of nuclei is hardly feasible because of their low brightness. Berman found that some nuclei showed a Planckian energy distribution in the visible spectrum with a temperature corresponding to the results of Zanstra's method. These were relatively low-temperature nuclei, with temperatures of about 30,000°. In cases when the allowed lines in the nebular spectrum give a higher nuclear temperature (as for Wolf—Rayet stars), the calculated temperature increases with increasing ionization potential of the particular atom. The temperatures estimated from the forbidden [O III] lines are close to the lower limit value, whereas the temperatures estimated by Ambartsumyan's method from the λ 4688 Å He II/H_β ratio are the highest. For example, for the nucleus of NGC 7027 this method gives 165,000°; other methods give much lower temperatures, down to 43,000°. The line intensities are also affected by a whole range of additional factors. However, no attempt has been made to summarize the situation and to find out which of the various temperatures are closest to the actual temperature. Therefore, the commonly used numerical data for the temperatures of the nuclei are in fact fairly arbitrary figures and their significance must not be overestimated.

The only certain thing is that the coldest nuclei are stars with O-type absorption spectra. Their temperatures are 25,000—30,000°. The temperatures of Wolf—Rayet nuclei are higher, generally reaching the normal values for stars of this type. The hottest nuclei are those with continuous spectra. The highest estimated temperature is that of the nucleus of NGC 6445, which is 170,000°.

As the temperature of the nucleus increases, the difference between its visible radiation and the radiation of the entire nebula becomes larger. It has been established that some nebulae without any visible nucleus definitely have a nucleus, which remains unobservable against the background of the high surface brightness of the nebula.

Distance determinations of the individual nebulae made it possible to estimate the absolute magnitudes of the nuclei. The photographic magnitudes of the nuclei were found to be much higher than those of normal

O-type and Wolf—Rayet stars and they showed a substantially greater dispersion, ranging from 0^m to $+7^m$, i.e., these are intermediate white, or more precisely blue, dwarfs. The hotter these nuclei, the lower their luminosities. The bolometric magnitudes show a smaller dispersion.

Their masses are not known, but the statistical method gives figures close to 2 solar masses. (The mass—luminosity curve gives masses close to 5.) Since the bolometric luminosity of all the nuclei is fairly constant, we come to the following conclusions.

The nuclei of planetary nebulae are a link in the white-blue sequence of the Hertzsprung—Russell diagram, lying between the normal Wolf—Rayet stars and the blue dwarfs. They are among the hottest known stars and their densities fall between those of Wolf—Rayet stars and white dwarfs. Their physical characteristics show a considerable spread.

The attempts to discover Einstein's gravitational red shift in the spectra of nuclei did not give conclusive results. In one case a shift of +54 km/sec was obtained, but other similar nuclei did not show any comparable effect. We do not know at what distance from the center of the star its spectral lines originate, either.

Swings and O. Struve established that the spectra of nuclei differ from the spectra of normal Wolf—Rayet stars, although at a first glance there is considerable likeness between them. The bright bands in the spectra of the nuclei studied were found to be narrower than those in the spectra of normal stars, and the carbon bands in spectra with nitrogen were stronger than those in normal nitrogen stars with carbon in their spectrum. These differences, however, are not very significant and may eventually prove to be quite accidental.

In conclusion note that the envelopes of planetary nebulae are genetically linked with their nuclei: they are most probably generated by the nuclei, although the exact mechanism of this effect is not clear. Shklovskii suggests that the nebulae are the result of dissipation of the atmospheres of some giant cold stars, e.g., RV Tau variables.

§ 76. Emission nebulae and hydrogen fields. Originally only discrete bright nebulae of diffuse shape were known, e.g., the Orion Nebula. Their spectra, besides bright emission lines, show a distinct continuum (Figure 127), partly of unknown origin, as the continuum of planetary nebulae, and partly representing the spectrum of light scattered by dust particles in the nebula. The continuous spectrum is often very strong, and the weak lines are indicatory of low ionization and low excitation of the nebular gas.

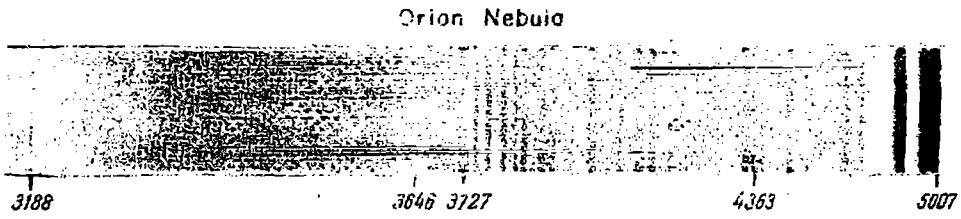


FIGURE 127

The nebular spectrograph developed at the Yerkes Observatory led to the discovery of new objects in this class. The nebular spectrograph reduces the light losses when photographing the spectra of extended and exceedingly weak luminous objects, while retaining all the advantages of a slit spectrograph (which, however, are only partly utilized). In the nebular spectrograph, the high-speed prism camera is placed directly after a long and narrow slit, and no collimator is used. This slit is situated far from the camera (virtually "at infinity" from the camera's viewpoint). Recently the slit has been replaced by a narrow mirror which rotates together with the camera tracking the celestial object, and the reflected light of the nebula hits a fixed plane mirror which is mounted up the scale at a distance of 20—40 m from the camera; the fixed mirror is tilted so that its normal is directed to the South Celestial Pole. The light reflected from this mirror reaches the prism camera.

Photographs taken with the nebular spectrograph revealed new weakly luminous nebulae and very extensive parts of the sky were found to be faintly luminous in hydrogen light and in the light of other typical nebular lines. These large extended areas were called hydrogen fields, mainly because the Balmer lines were the strongest in their spectra. No distinct limits have been fixed between weak nebulae and hydrogen fields.

Hydrogen fields are more easily detected on photographs taken with high-speed cameras through a red filter, which transmits the H_α hydrogen line but blocks out almost completely the nearby continuum and the auroral red lines. Under these conditions, long exposures are possible without the danger of fogging. This method was applied at the Crimean Astrophysical Observatory by Shain and Gaze, and later by other observers as well. Their plates revealed a wealth of fine features in previously known nebulae and also showed new nebulae and hydrogen fields.

It is impossible to say how many gaseous nebulae are known. Their classification is most disorderly, and one nebula often has more than ten different numbers, assigned to various visible features in it. Some nebulae are observed against the background of extended hydrogen fields, where each condensation and feature receives a separate number, irrespective of the number assigned to the field as such. The number of items in Cederblad's catalogue, in the lists of the Crimean Observatory, and in other lists is close to one thousand.

Statistical investigations of diffuse gaseous nebulae therefore rest on very shaky ground, unless we limit our sample to the brightest nebulae and carefully chart their limits irrespective of the enumeration system used in the catalogues.

Using the brightest nebulae, Hubble established in 1921 that an O-type star or a B0 star is generally observed inside the gaseous nebulae or near them. The glow of the diffuse gaseous nebulae is thus apparently excited by these very hot stars. Note, however, that the exceedingly hot Wolf—Rayet stars are seldom associated with gaseous nebulae. If stars of late B subtypes are observed in association with diffuse nebulae, this probably means that the spectral type of the star has been incorrectly determined. In certain cases, however, the nebulae are actually associated with B1 stars, and then the spectrum of the nebula is characterized by low excitation and low ionization, and sometimes the nebula shows a continuous spectrum produced by a scattering dust component.

Occasionally, small nebulae are clearly linked with late-type stars, whose spectra invariably show bright high-excitation lines. In these cases the nebula is apparently excited by invisible hot companions of the cold stars or by nova-like companions. An interesting example is provided by the nebula surrounding the red variable star R Aqr.

The assumption that the diffuse nebulae, like planetary nebulae, are excited by nearby hot stars was verified by correlating the brightness of these stars with the diameter of the associated nebulae under conditions of equal exposure.

If we assume that the visible light of a star of apparent magnitude m is not absorbed in the nebula and that the brightness at any point in the nebula is inversely proportional to the square of the distance of that point from the star, the range a_1 over which the nebula can be traced for a given exposure is theoretically expressed by the equality

$$m + 5 \lg a_1 = \text{const.}$$

Observations show a good fit with this formula; hence the conclusion that the nebular glow is indeed excited by the star.

Later it was established that the emission of allowed lines by the nebula was due to photoionization followed by recombination. The emission intensity in the hydrogen line should increase with the temperature of the exciting star, since the number of the ionizing ultraviolet quanta emitted by the star increases with increasing temperature. Having reached this conclusion, Zanstra determined the temperatures of the exciting stars from the brightness of the nebulae and obtained a good fit with the known temperatures of B0 and O stars.

However, there are a number of nebulae, filamentary nebulae in particular, which seem to have no exciting star. Their emission is possibly attributed to high-velocity collision of gas or dust clouds. This collision produces a compression shock, and the temperature rapidly increases (Oort's hypothesis). The emission is sometimes attributed to electromagnetic effects.

The hydrogen fields present a more complex problem than the discrete nebulae. For example, the large circular weak nebula around λ Ori is regarded as a hydrogen field 70 pc in diameter. Formations 30 to 70 pc in diameter are generally classified as hydrogen fields, to distinguish them from discrete bright nebulae whose diameter seldom exceed a few parsecs. Both the discrete emission nebulae and the hydrogen fields are sometimes referred to as HII regions, to distinguish them from the nonluminous HI regions of neutral hydrogen.

The hydrogen fields in Cepheus, Cassiopeia, Cygnus, and Sagittarius are so large, however, that it is impossible to identify the exciting stars. Their spatial extent and dimensions therefore cannot be established either.

It is not clear why whole clusters of hot stars, e.g., in Orion and Sagittarius, illuminate much smaller nebulae than a single star λ Ori and why the glow of the nebula around λ Ori is weak and almost uniform over the entire surface. This nebula is probably a plane layer of gas, and not a gas-filled volume.

The gaseous nebulae and hydrogen fields show a pronounced concentration to the plane of the galactic equator, where the exciting hot stars concentrate. Moreover, these hot stars in other galaxies are concentrated in spiral arms, and it is in the arms that the gaseous nebulae should be looked for.

Indeed, M 33 in Triangulum contains several tens of gaseous nebulae and in M 31 in Andromeda Baade discovered over 300 gaseous nebulae. Initially the brightest nebulae were detected in the spectra obtained with a slit spectrograph, and later in the spectra taken with an objective prism. The next stage was the search for gaseous nebulae on photographs shot through a narrow-band H_α filter. Of all the spectral lines of diffuse nebulae, H_α produces the maximum response in panchromatic plates.

In these spiral galaxies the nebulae or the hydrogen fields are concentrated in the spiral arms, sometimes very far from their nuclei. Their radial velocities make it possible to determine the law of galactic rotation at large distances from the center.

A few dozen gaseous objects were observed in the Magellanic Clouds. Of these, 30 Dor has the brightness record. A number of nebulae in other nearby irregular galaxies were discovered and studied at the Crimean Observatory.

In more distant galaxies, the presence of emitting gas is detected only from the total spectrum. The only indication of interstellar gas is the λ 3727 Å [O II] line, which is exceedingly strong in low-excitation nebulae and is generally superimposed on the weak ultraviolet end of the galacitc continuum. The H emission line is more difficult to detect in these spectra, and the two lines are not always present simultaneously. This points to certain differences in the predominant gas emission in different galaxies.

An even more significant feature is that emission lines are observed in the spectra of 80% of the late-type spirals Sc, Sb and especially in irregular galaxies, i.e., galaxies containing numerous hot stars and rich in interstellar dust. Fewer emission lines are encountered in early-type spirals, and only 20% of elliptical galaxies (E) show bright spectral lines.

Since diffuse nebulae are on the whole associated with colder stars than the nuclei of planetary nebulae, their spectra show a lower degree of ionization and excitation. As a rule, the [O III] "nebulium" line in their spectra is weaker than H_β, at variance with the spectra of most planetary nebulae. The strongest lines in the spectra of diffuse nebulae are [O II], and not [O III]. The excitation and ionization of "hydrogen fields" is even lower. Here hydrogen lines are dominant, although [O II] lines are observed and there are also strong [O I] lines, which are characteristic of the auroral spectra.

The spectra of diffuse gaseous matter provide valuable information on the physical state of interstellar gas. For instance, nebulae excited by the collisional mechanism should have much stronger forbidden lines than nebulae fluorescing following excitation by the ultraviolet radiation of the nucleus star.

The spectra of diffuse nebulae and fields in the Galaxy have been poorly studied, despite the invention of nebular spectrographs. The Orion Nebula, the Cygnus Loop, and the North America Nebula are the only ones which have been studied in any detail. The radial velocities, for example, are known only for about a dozen nebulae.

In the unusually bright Orion Nebula, the radial velocities of different parts of the nebula were measured and semibiased local differences reaching 12 km/sec were detected. It is difficult to interpret these differences, since the observed radial velocity is the average for entire

gas column in the direction of line of sight. Different authors variously attribute these results to rotation, expansion, or large-scale turbulence. Small-scale turbulence is responsible for smaller differences in radial velocities and for Doppler broadening of lines. In this respect, highly interesting results were provided by interferometric observations of the nebula in monochromatic light.

Large-scale turbulence in nebulae is manifested in the form of curls, filaments, veinlets, and other features generally associated with turbulent flow.

The onset of turbulence is generally determined by the Reynolds number

$$R = \frac{\varrho v l}{\eta}$$

where l is the diameter of the object, v are the relative velocities, ϱ is the density, and η is the viscosity ($\nu = \eta/\varrho$ is the kinematic viscosity). For nebulae 10 pc in diameter with kinematic viscosity of 10^{19} cm^2/sec, R is close to 10^6. Laminar flow gives way to turbulent flow for $R > 1000$, so that nebulae are evidently gas formations in a state of turbulence.

If v_h is the mean eddy velocity on a scale less than λ_h $\left(k = \frac{2\pi}{\lambda}\right)$, we have according to Kolmogorov

$$v_h = \text{const}\, \lambda^{\frac{1}{3}}.$$

The Kolmogorov eddy spectrum can be written in the form

$$[(v_1 - v_2)^2] = c^2 |r_1 - r_2|^{\frac{2}{3}},$$

where $v_1 - v_2$ is the difference in radial velocities between the points r_1 and r_2. Using this spectrum, we can study the turbulence of nebulae from observations.

For the Orion Nebula the n in the equality $v_h = \text{const}\, \lambda^n$ was found to lie between 1/4 and 1/2 and the smallest eddy elements are larger than 0.02 pc.

Instead of searching for correlation between velocity differences and the separation of pairs of points, one can concentrate on surface brightness fluctuations in the nebula, assuming that they are entirely attributable to variations in the density of the emitting gas. In this way Aller found that the largest eddies in the Cygnus hydrogen field were up to 10 pc in diameter, and their energy was gradually transferred to smaller eddies. However, Aller himself noted that brightness fluctuations may be due to other reasons as well. Apart from turbulence, they are affected by gravitation, magnetic fields, and various other factors. Collisional processes may give rise to shock waves. The electron temperature and the gas density in nebulae can be estimated from the intensity of spectral lines using the theory of physical processes in nebulae, described in Chapter XVI.

The electron temperatures in gaseous nebulae are of the order of 10,000°, i.e., they are not markedly different from the values obtained for planetary nebulae. The densities are on the average lower than in planetary nebulae: they range from a few atoms to a few tens of atoms in 1 cm^3.

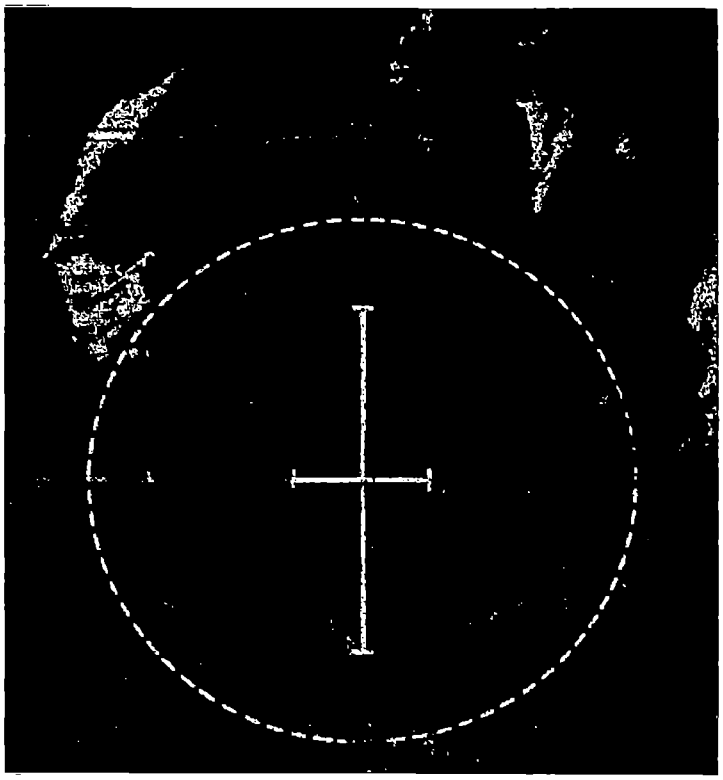


FIGURE 128

Most, if not all, diffuse nebulae are mere condensations in the general substratum of interstellar gas. This, however, does not mean that the nebulae are the result of compression of the gas medium. The converse is probably true: the interstellar medium is steadily replenished by gas from dissipating nebulae. The magnetic fields in nebulae probably slow down their dissipation. The magnetic field geometry is apparently responsible for the peculiar structure of the filamentary nebulae (Figure 128). Gaseous nebulae are genetically linked with the exciting hot stars (at least this is so in most cases). The best proof of this is that the velocities of some nebulae differ only by a few km/sec from the velocities of the associated stars. Furthermore, nebulae illuminated by hot stars are more numerous than what follows from the theory of chance encounters. The star often occupies a special position relative to the nebula. This factor is particularly pronounced for what Shain called peripheral nebulae (Figure 129). These are not unlike giant planetary nebulae with a ring effect. Their

brightness is higher at the periphery, and the exciting star, or sometimes a whole cluster of stars, is situated at the center of the ring. Peripheral nebulae always show distinct signs of outward radial flow of gas.

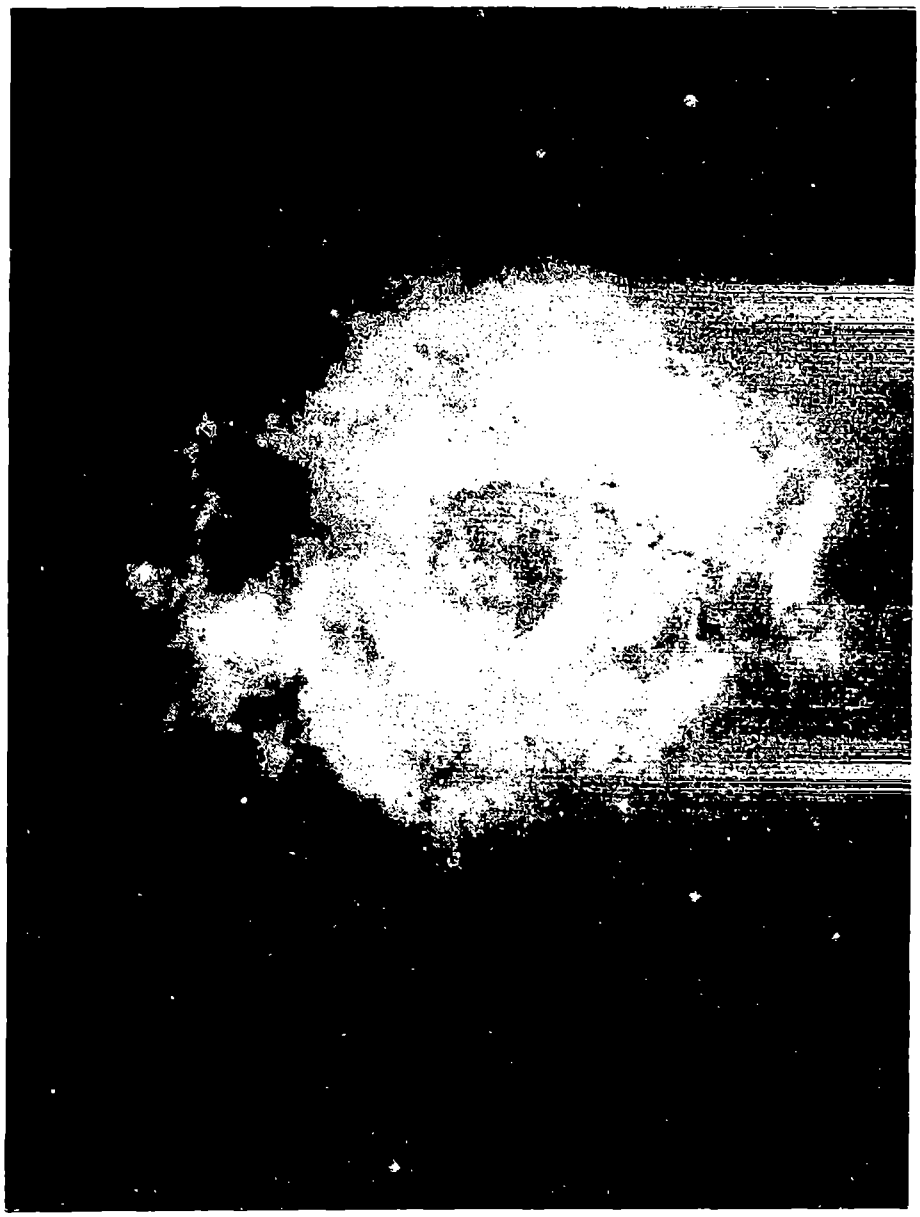


FIGURE 129

Filamentary nebulae, conversely, have no obvious exciting stars. The filaments are often parallel to one another. In the Cygnus Loop (NGC 6960) the filaments are particularly thin and narrow. Most of them are no wider than 2", and their linear width is probably a few thousand astronomical units. (Note that the distances to these nebulae cannot be measured, as they are not associated with any stars of known luminosity).

Numerous hydrogen fields and nebulae are foggy formations without any apparent structure.

Some nebulae are comet-like or fan-shaped (NGC 2261, Figure 130). The exciting star is situated in the head or outside the nebula near its tip. These "comets" are often associated with T Tau stars. Some of these nebulae have variable brightness. Patchy nebulae with relatively sharp edges are fairly frequent. Nebulae in the form of narrow strips are also observed.

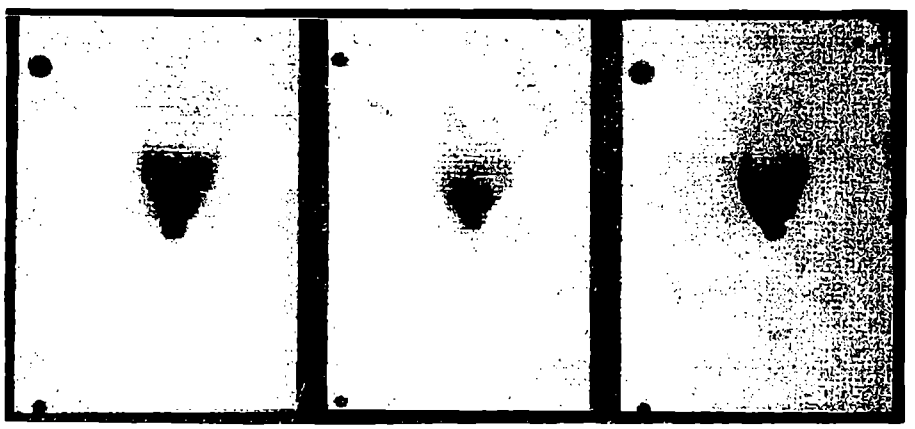


FIGURE 130

No uniform morphological classification of diffuse nebulae has been developed.

More or less reliable mass estimates of diffuse gaseous nebulae were obtained quite recently using the theory of nebular emission. Ambartsumyan's method was used, in particular. Mass determinations carried out in Moscow and in Simeise lead to the following order of magnitude estimates (masses in units of solar mass):

LARGE DISCRETE AND BRIGHT NEBULAE

Orion Nebula (gaseous component)	150
NGC 2237 (Large Nebulae in Monocerps)	260
Omega Nebula in Sagittarius	500
NGC 6523 in Sagittarius	3200

Other galaxies contain occasional giant nebulae. The brightest and the largest is 30 Dor in the Large Magellanic Cloud. Shain estimated the

masses of some nebulae in the spiral galaxy in Triangulum and obtained figures of up to $10^4 M_\odot$. These giant nebulae are probably clusters of nebulae, rather than discrete objects. At the other extreme of the scale, multitudes of weak and small nebulae should have masses of the order of a few solar masses and even fractions of the solar mass. The highly tenuous but extensive hydrogen fields may have very large masses. The masses of gaseous nebulae thus range between wide limits. Their densities are also greatly different. The density at the center of the Orion Nebula is of the order of 10^{-20} g/cm^3 (10^4 hydrogen atoms in 1 cm^3), in other nebulae it drops to 10^{-22} g/cm^3, and in the space between the nebulae the density is 10^{-24} g/cm^3.

§77. Interstellar gas. The existence of discrete gaseous nebulae was established long ago, but the existence of the general interstellar gaseous medium was a matter of gradual proof, which began back in 1904. In that year Hartmann found sharp and narrow H and K lines of ionized calcium in the spectra of hot spectroscopic binaries, which did not share the shifts of the other lines in the course of the orbital motion of the components. These static calcium lines were originally attributed to a cloud of Ca^+ vapor in which the two components were immersed. Later the corresponding line intensities were found to increase for the more distant stars and for stars closer to the galactic plane. It thus became clear that the emitting gas was not concentrated near certain stars but actually filled the entire interstellar space, its concentration increasing toward the galactic plane.

In the spectra of distant cold stars the interstellar Ca^+ lines are obscured by the strong lines of this ion originating in the stellar atmospheres, whereas the atmospheres of hot stars contain no Ca^+: calcium is present there in higher ionization stages. The line intensity of interstellar Ca^+ increases with increasing distance to the star, since the optical thickness of the Ca^+ gas between the Sun and the star increases. If we find the correlation between the line intensity and the distance, we will be able to estimate unknown distances from measurements of interstellar line intensities. The accuracy of this method is not very high, since the distribution of Ca^+ in space is inhomogeneous.

Interstellar calcium takes part in the galacitc rotation. The mean shift of its lines due to the differential rotation effect is approximately equal to half the shift of these lines in stellar spectra. The galactic differential rotation effect is proprotional to the distance from the Sun, and the distance to the center of mass of an interstellar Ca^+ column between the Sun and any given star is approximately equal to half the distance to the star.

Interstellar Na lines were discovered in 1919, and interstellar lines of K, Fe, Ti, and neutral calcium were discovered after 1937. Narrow absorption bands of interstellar CH, CH^+, and CN and a number of molecular bands of unknown origin were also observed. The strongest of these is the 4431 Å band.

The discovery of these lines established the existence of H, C, and N in interstellar space.

A characteristic property of all the observed interstellar lines and bands is that they are produced by transitions from the lowermost level and even sublevel. In molecules light-absorbing transitions originate on the lowermost vibrational, rotational, and electronic levels. Therefore only some of the lines characteristic of the corresponding molecular bands are observed and they are unusually narrow.

The extremely low density of the interstellar gas and the great distances from stars are responsible for the very low frequency of collisions of atoms and molecules with other particles and light quanta. The great majority of the atoms and molecules are therefore in the unexcited ground state. When and if a quantum is absorbed, it is thus absorbed by an atom or a molecule in a state of minimum energy.

This explains why the omnipresent hydrogen makes no contribution to interstellar absorption of light. The transition of atomic hydrogen from ground state to an excited state involves absorption of quanta in the Lyman series lines. The Lyman lines, however, lie in the far ultraviolet and are thus inaccessible to optical observations. Other atoms, though possibly present in the interstellar space together with hydrogen, are not detected either if the lines of their principal series lie in the inaccessible ultraviolet.

The sharpness of the interstellar lines compared to the lines originating in stellar atmospheres is attributed to the low number of absorbing atoms. Doppler effect due to thermal or turbulent motions in the interstellar medium is the only factor which perceptibly broadens the spectral lines.

The kinetic temperature of the interstellar gas depends on the temperature of the exciting stars and on the degree of ionization (which is again a function of the star temperature). Originally, the temperature of the interstellar gas was set equal to the temperature of the exciting stars, i.e., around 10,000°. Further analysis has shown, however, that this is true only for regions where hydrogen (the main interstellar gas component, although it does not show in the absorption spectrum) is completely ionized. In regions of neutral hydrogen, collisions of hydrogen atoms with dust particles and other atoms result in a loss of kinetic energy (due to collisional excitation of the other particles) and lower the gas temperature. The temperature of the interstellar medium containing neutral hydrogen is around 100°K; this result is confirmed by radio observations (see below).

The interstellar gas is ionized by the high-frequency quanta of stellar radiation. This radiation is diluted; however, introduction of the dilution factor (of the order of 10^{-16}) in the ionization equation is insufficient. Indeed, the resultant radiation from various stars if far from being Planckian. The problem is further complicated by the fact that we do not know the frequency of stars of different temperatures and the luminosity distribution outside a small neighborhood of the Sun, and integration of energy over the spectrum in distant regions of the Galaxy is therefore highly uncertain. The gas ionization will vary between wide limits with the luminosity function of stars in different parts of the Galaxy.

Even for the immediate neighborhood of the Sun, the equation for the degree of ionization as a function of star temperature and kinetic temperature cannot be used without correction coefficients. At 100°K these correction coefficients for Na I, Ca I, Ca II, and K I range from 10 to 1600 and lower the ionization estimates.

Furthermore, the ionization equation always contains n_e, the number of free electrons in $1\,cm^3$, since these electrons govern the recombination processes which balance the ionization events. The effect of n_e is different for different atoms: it is extremely small for Ti II and very large for Ca I, say.

Determination of n_e from observations has not given reliable results so far. It is generally obtained by comparing the densities of Ca I and Ca II determined from the equivalent widths of their lines, but λ 4226 Å Ca I is observed only when the H and K lines are strong and "saturated," i.e., when their equivalent width is no longer sensitive to the optical thickness of the gas layer at the corresponding wavelengths. Other methods also fail to provide accurate data. The inaccuracy is also associated with the great inhomogeneity of the interstellar medium in various directions and various distances. The estimates apply to bright stars with strong absorption lines, i.e., they apparently correspond only to relatively dense gas clouds. The application of the improved theory of ionization, mentioned in the preceding, gives crude estimates for the relative abundance of some elements. Strömgren's abundance data (1949) are the following:

	H	Na	Ca	K	Ti
Sun's atmosphere	8	1	1.4	0.08	0.05
Interstellar medium in the direction of χ^2 Ori	2	1	2	0.06	0.03

The absolute figures for this gas cloud are 40 atoms of hydrogen in $1\,cm^3$, $n_H = n_e = 1000$.

In the space between these dense clouds the gas is much more tenuous, but the quantitative data are highly uncertain since we are dealing with extremely weak Na I and Ca II lines (the other lines being undetectable under these conditions). The theory cannot give a reliable value for the ratio n_{CaI}/n_{CaII}. The concentration of hydrogen is estimated at being less than 1 atom in $1\,cm^3$. In medium density clouds the concentration reaches about 1 atom in $1\,cm^3$. The density of the spectroscopically undetectable hydrogen is estimated with great uncertainty. Since this hydrogen is neutral, the free electrons are mainly supplied by Ca I, whereas the electrons released in the ionization of Ca II and other elements are few.

The determination of the quantitative chemical composition of the interstellar gas meets with additional difficulties. High-dispersion spectra of some stars show interstellar Ca II multiplets (Figure 131). The multiplicity of these lines evidently indicates that there are several interstellar gas clouds along the line of sight, all moving with different velocities. These results in fact provided a demonstration of the patchy, cloud-like structure of the interstellar gas. The spectra of 95 out of 300 fairly distant bright stars show double Ca II lines, 17 spectra contain triplets, and 4 spectra quadruplets. The peculiar velocities of the clouds are at most 5 km/sec. The average velocity according to Mel'nikov is 7 km/sec. However, 21 out of 166 stars show components with shifts exceeding 30 km/sec, and 7 stars have shifts greater than 50 km/sec; in one cloud the shift reaches 96 km/sec.

In some cases, the Ca II doublets are accompanied by Na I singlets. This means that the sodium in one of the clouds is ionized or absent altogether.

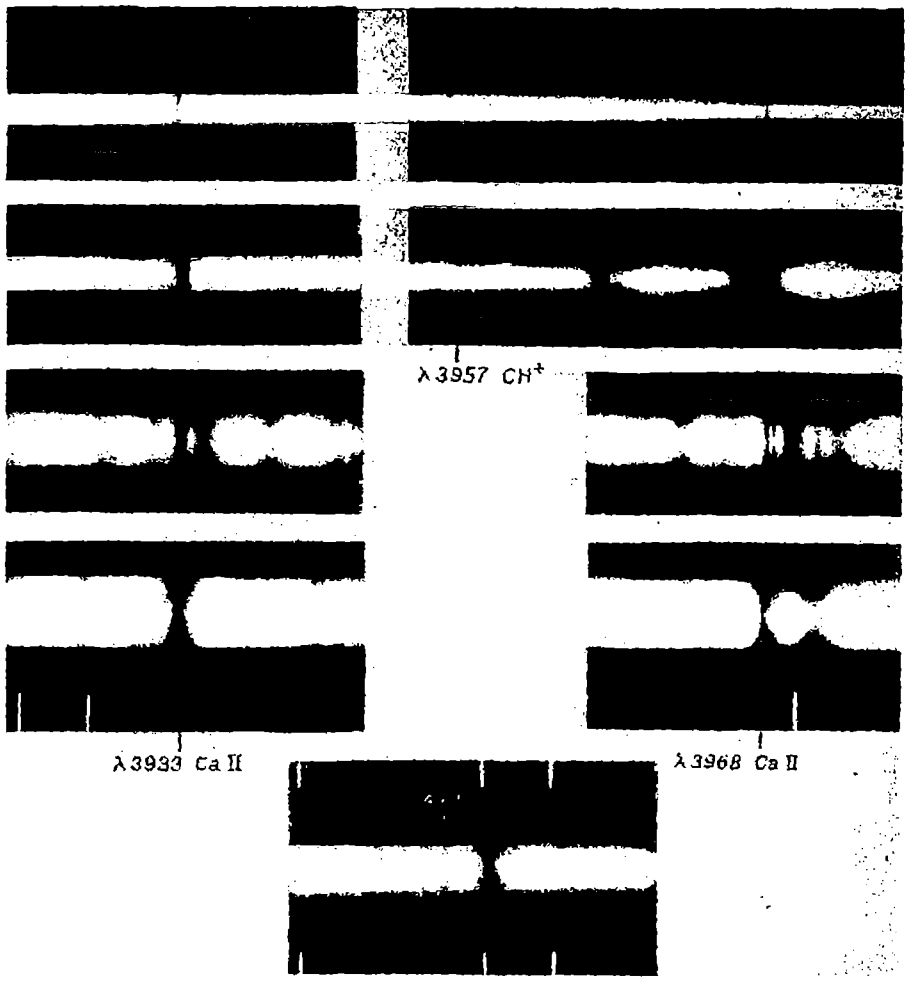


FIGURE 131

Some clouds show relatively strong molecular lines, others have strong metal lines. In HD 190603 the shift of the CH^+ lines is equal to that of one of the K components. The CH lines in the spectrum of χ Aur show show the shift as one of the K components, and the shift of the broadened CH^+ lines is equal to the average shift of the other K components. The CH/CH^+ intensity ratio varies from point to point.

The dimensions of interstellar gas clouds are apparently a few parsecs on the average. Their differential motions are partly attributed to galactic rotation and partly to turbulence, which is apparently quite considerable.

In regions of neutral hydrogen, at densities of 10 atoms in 1 cm^3 and temperatures of 50°K, the velocity of sound should be 0.8 km/sec; the free

path of hydrogen atoms is then 20 a.u. and of electrons $3 \cdot 10^{-3}$ a.u. The interstellar hydrogen is opaque in its 21 cm emission line, and radio observations can be applied to measure its temperature; they give 125°K in the direction of the galactic center.

Polarization of the light of distant stars was first discovered in 1949 by Hiltner and Hall and independently in 1950 by Dombrovskii. It reaches 10% and is almost independent of wavelength. The plane of oscillation of the electric vector in most parts of the Milky Way is close to the plane of the galactic equator. This effect is most prominent in the direction where the spiral arms presumably meet the line of sight. The polarization of the light of distant stars, while the near stars remain unpolarized, and the similar orientation of the polarization vector for different stars from one neighborhood suggested that the interstellar medium was responsible for this polarization. The overall increase of the percentage polarization with increasing interstellar absorption for the same stars corroborates this conclusion; polarization of star light is thus assigned to nonuniform absorption of light by elongated dust particles with variously oriented electric vector. When the light is incident in the direction of the long axis of the particle, the absorption is stronger. This indicates that the dust particles are aligned roughly at right angles to the spiral arms. The situation, of course, is much more complex when the line of sight intercepts several dust clouds of inhomogeneous structure.

It was also suggested that the dust particles are rich in iron and are therefore aligned by the magnetic field. This is an improbable assumption: the iron content of the dust particles cannot be high, but they are paramagnetic and are thus aligned with the major axis at right angles to the magnetic field.

Thus, the structure of some bright nebulae and the polarization of light in interstellar space point to the existence of galactic magnetic fields of $10^{-4} - 10^{-5}$ gauss. The radius of curvature of a charged particle in these fields is therefore 10^{-7} of its free path. The interstellar gas thus largely retains the ideal gas properties, but the application of the laws of hydrodynamics to motion at right angles to the magnetic field is questionable. It is by no means certain that the magnetic field is "frozen" into the interstellar gas, since the small radii of curvature of the free electrons lower the electrical conductivity of the gas and reduce the relaxation time.

The latter point is of considerable significance for the interpretation filamentary nebulae. In neutral hydrogen regions, pressure differences are smoothed out through propagation of elastic compression waves, which travel with the velocity of sound, i.e., 0.8 km/sec or 1 pc in 10^6 years. Pressure differences are thus maintained on the cosmic time scale, giving rise to interstellar winds.

On the other hand, calculations show that a low-density neutral hydrogen heated during an accidental encounter with hot stars cools down in 10^9 years. In small high-density clouds the gas temperature is low. Cold clouds are heavier and thus fall to regions of lower potential energy, i.e., toward the galactic plane. As they fall, they displace hot gas masses which rise to higher galactic latitudes. This mechanism is sufficient to produce turbulence.

The diffuse interstellar matter is linked, at least partly, with ejection of stellar matter. The active stars are distributed nonuniformly in the Galaxy, and

for this reason some regions acquire relatively high gas density and pressure. The high-pressure gas expands and colliding with older clouds it produces turbulence. The turbulence and motion of interstellar gas is associated with a variety of reasons; without them, however, the observed inhomogeneities and cloud motions could hardly persist.

The effect of radiation pressure on the motion of interstellar gas has not been elucidated. It is clear, however, that the gas should experience a certain drag due to the motion of dust particles which are definitely impelled by the radiation pressure of the nearby stars.

§ 78. HI and HII regions. What is the relationship between the invisible clouds of interstellar gas, detected only by their absorption lines, and the bright gas clouds reminiscent of nebulae and hydrogen fields? It seems that the bright discrete nebulae are mostly mere condensations in the interstellar gas medium, distinguished by their higher density and stronger emission. These distinctive features are almost invariably associated with the presence of hot stars or clusters of stars in or near the nebula; there is apparently a definite genetic link between these stars and the nebula. We cannot say whether or not similar discrete and dense clouds exist which remain invisible for lack of hot exciting stars. On the other hand, extensive areas with multitudes of hot stars, e.g., Perseus, are entirely free from luminous gaseous matter and are thus absolutely gas-free.

Weak luminous regions, in particular hydrogen fields, are probably fundamentally identical to gas clouds detected by their absorption lines. The only difference between these two groups of objects is that the latter are not associated with any hot stars and therefore remain dark and nonluminous. Hydrogen emits through photoionization followed by recombination. In luminous clouds hydrogen is thus almost completely ionized, whereas neutral hydrogen clouds neither emit nor absorb. Interstellar gas is correspondingly divided into HI and HII regions. Regions of both types contain other gases as well, but their physical conditions are uniquely related to the state of hydrogen in them.

The theory of HII regions was developed by Strömgren. Near a hot star, the energy beyond the Lyman limit (912 Å) is sufficient for complete ionization of the surrounding hydrogen. The radiation dilution coefficient $W=R_*^2/4r^2$ indicates that the radiation is attenuated with distance from the star (here R_* is the radius of the star and r is the distance from the star). The rate of ionization decreases in proportion to W, whereas the number of collisional recombinations does not change with increasing distance from the star. The number of neutral atoms thus increases with increasing r. Neutral hydrogen readily absorbs the quanta with $\lambda < 912$ Å, whereas ionized hydrogen is transparent to these frequencies. A sharp spherical boundary is thus formed around the star, beyond which the hydrogen is not ionized and the quanta with $\lambda < 912$ Å do not pass.

The degree of ionization of hydrogen is described by Saha's modified equation

$$\frac{n_i n_e}{n_0} = \left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} \frac{2g_i}{g_0} 10^{-\frac{5040}{T}13.54} \left(\frac{T_e}{T}\right)^{\frac{1}{2}} We^{-\tau_u}.$$

Here n_i is the number of protons, n_0 is the number of neutral atoms, g are the statistical weights of the respective states, T_e is the electron temperature, T is the temperature of the star, τ_u is the opacity of the gas at

$\lambda < 912$ Å; 13.54 is the ionization potential of hydrogen. Taking $n_i + n_0 = n$ and writing $x = n_i / n$ for the degree of ionization, we have

$$\frac{n_i n_e}{n_0} = \frac{n x^2}{1-x} = \frac{c}{r^2} e^{-\tau_u},$$

where c incorporates all the constants of the previous equation. The absorption coefficient near the Lyman limit is $\sim 10^{-17}$ cm^2 per atom, and the opacity of hydrogen is

$$d\tau_u = (1-x)\, n a_u 3 \cdot 10^{18}\, dr.$$

Inserting this expression in the above equality, we obtain

$$e^{-\tau_u}\, d\tau_u = \frac{3 \cdot 10^{18}}{c}\, a_u n^2 c^2 r^2\, dr.$$

To solve this differential equation we use the following substitution of variables:

$$y = e^{-\tau_u}, \qquad z = \frac{3 \cdot 10^{18} a_u n^2 r^3}{3c}. \qquad \alpha = \left[\frac{9}{nc\,(3 \cdot 10^{18} a_u)^2}\right]^{\frac{1}{3}}.$$

Then

$$\frac{dy}{dz} = -x^2,$$

$$\frac{1-x}{x^2} = \frac{\alpha z^{2/3}}{y}.$$

Since $3 \cdot 10^{18}\ a_u \approx 10$ and $n \approx 1$, we have $\alpha \approx c^{-1/3} \approx R^{-2/3}$, which is small in units of the solar radius. For $y = 1$ the hydrogen is ionized and thus transparent. In this case $x = 1$ and $y = 1 - z$. As r increases, z slowly increases approaching unity, while y goes to zero and the ionization falls off rapidly. The critical radius r_0 is obtained for $z = 1$, i.e.,

$$r_0 = \left(\frac{c}{10^8 a_u n^2}\right)^{\frac{1}{3}}.$$

Strömgren calculated for $\alpha = 0.01$ the values of the relevant parameters when r/r_0 goes through 1.

The radius of complete ionization r_0 depends on $n^{-2/3}$ and on the temperature of the star. Numerical data, however, are highly uncertain because of divergences in temperatures of the hottest stars. Strömgren found that the volume of complete ionization around a O8 star is 216 times greater than that around a B3 star and 26 times greater than around a B0 star. If, however, the temperature differences between stars of these types are smaller, the ratio of respective volumes for O5 and B0 stars will diminish to a mere 2.5. In a gaseous medium with one H atom per cm^3, the radius of complete ionization for an O star is not less than 30 pc, possibly reaching 100 pc; between the clouds, where $n = 0.1$ or 0.001, the radius of complete ionization is even larger. The gigantic hydrogen clouds extending over hundreds of parsecs may thus be excited by very few stars. The absence of He I emission in H II regions is not clear at this stage.

Some authors also treated planetary nebulae as spheres of completely ionized hydrogen in much larger gaseous clouds, but this approach is apparently erroneous.

The sum total of HII regions near the Sun is about 1/10 of the HI regions. The electron temperatures in HII regions, as in nebulae, are about 10,000°, i.e., much higher than in HI regions, whereas the average densities are probably the same. The velocity of sound in HII regions is about 17 km/sec (for $n = 1$), and the pressure differences are thus smoothed out much faster than in HI regions. The radius of curvature of the trajectories of charged particles in HII regions is about 100 times as large as that in HI regions, and this markedly affects the motion of gas in magnetic fields. Hot HII regions should exhibit a tendency to rise above the galactic plane.

Note that according to Pikel'ner highly tenuous interstellar gas in the same state as in HI regions in the galactic layer also occurs at great distances from the plane of the Galaxy. The velocities of this gas may be quite substantial.

In view of the preceding, no reliable estimate of the mass of the gas component of the Galaxy can be obtained. Dynamic considerations based on a comparison of the theory of galactic rotation with observations show that the mass of the diffuse matter near the Sun is comparable with the total mass of stars in that volume. This is an upper bound estimate. The various attempts to differentiate between the proportion of gas and dust are most unconvincing. Radio observations give a more precise estimate of the mass of the gas component of the Galaxy.

§ 79. Radio observations of the gaseous diffuse matter. Radio observations of interstellar gas are quite a recent development. Interstellar ionized hydrogen gives off thermal radio emission whose intensity for an optically thin gas layer is independent of frequency. At waves shorter than 1 m, this is the dominant component of the radio spectrum. The radio emission of interstellar gas has different intensities in different directions. It is markedly concentrated toward the plane of the Galaxy and its center. Shklovskii managed to isolate this component from the general radio flux density from outer space. Besides the principal maximum in Sagittarius, it has two secondary maxima in Cygnus and Perseus. The connection with the concentration of gas in the spiral arms of the galaxy is not clear, since relatively near gas masses may give under certain conditions a larger apparent effect than bigger but more distant masses. Anyhow, the radio emission of ionized hydrogen points to the presence of hot stars ionizing the interstellar gas. Ionized hydrogen absorbs long radio waves and is thus highly opaque in bright nebulae at wavelengths over 5—7 m and in weak nebulae at wavelengths over 10—15 m.

Besides the continuous radio spectrum of ionized hydrogen, van de Hulst and Shklovskii predicted monochromatic radio emission at 21 cm wavelength from neutral hydrogen. This radiation is emitted in transitions between the two hyperfine structure components of atomic hydrogen. The intensity of this radiation also gives very low figures for the temperature of neutral hydrogen regions, about 25—50° and 125°K in the direction to the center of the Galaxy. Because of this low temperature, the interstellar neutral hydrogen is opaque at $\lambda = 21$ cm. However, the differential effect of galactic rotation in radial velocities produces a noticeable Doppler shift in various directions, so that the gas moving ralative to the Sun with a certain velocity emits radio waves whose wavelength is slightly different from 21 cm.

These waves are freely transmitted by the interstellar hydrogen, so that the distribution of neutral hydrogen in the Galaxy can be charted over enormous distances from superimposed lines of various gas clouds moving with different velocities in the Galaxy. This method enables us to penetrate over thousands of parsecs into the interior of the Galaxy, but the interpretation of the observations is based on various highly simplified models of galactic structure and rotation and is therefore highly unreliable (Figure 132). For example, the spiral arms of the Galaxy charted from the observed condensations of neutral hydrogen have very little in common with the spiral arms of other galaxies.

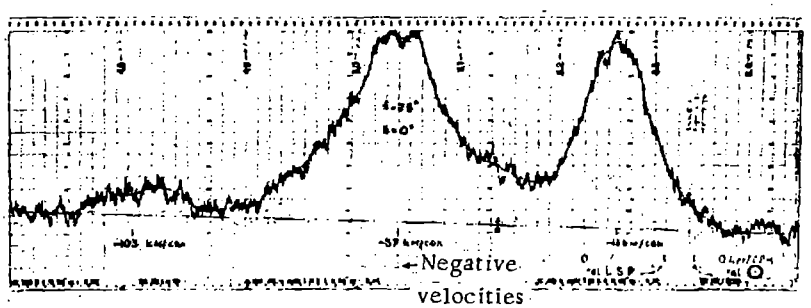


FIGURE 132

Interstellar molecules also should emit monochromatic radio waves. Entirely new horizons are thus opened for the survey of diffuse gaseous matter in the Galaxy.

Apart from the thermal radio waves emitted by the diffuse gas in the Galaxy, there is another radio component which prevails at low frequencies and whose intensity falls off with increasing frequency. This component is produced by discrete radio sources which do not show any apparent concentration toward the galactic plane and reveal only a slight concentration toward the center of the Galaxy. These sources are sometimes called radio stars. However, application of high resolution radio telescopes* has shown that some of these radio stars have measurable (though small) angular diameters. Their radio emission is generally attributed to the bremsstrahlung of relativistic electrons in magnetic fields of some $10^{-5} - 10^{-6}$ gauss, which apparently fill the interstellar space.

It has been definitely established that the so-called radio stars are of partly galactic and partly metagalactic origin. The radio sources should be identifiable with ordinary, optical gaseous nebulae, and such identifications were indeed made in some cases. Nebulae which are remnants of supernova explosions, in particular the Crab Nebula in Taurus, are also powerful

* A common radio telescope comprises a radio receiver mounted at the focus of a concave metallic reflector which can be pointed to different parts of the sky. The resolving power of the radio telescope increases with increasing reflector diameter, but there are obvious technical limitations to the size of radio telescopes. Another type of radio telescope is a moving or partly moving array of antennas, which gives a higher resolving power but is limited in other respects.

radio sources. Their radio intensity is frequency independent, whereas the intensity of metagalactic radio sources is proportional to the wavelength. The radio emission of supernova remnants is apparently not associated with the visible optical light of the nebula: it seems to be the result of bremsstrahlung of relativistic electrons. The radio flux density of Crab Nebula at 20 cm is about $2 \cdot 10^{-23}$ watt/MHz, and the flux density of the supernova remnant in Cassiopeia (the most powerful of known radio sources) is double this figure.

Other powerful discrete radio sources are peculiar galaxies with strong hydrogen emission lines in their spectra. Some of these galaxies are spherical. Others, although spherical, are bisected by a dark lane along the middle (probably a concentration of dust nebulae). The radio emission of the Magellanic Clouds has been studied in some detail; the radio velocities of neutral hydrogen in the Clouds made it possible to determine their rotation and mass.

Weak radio emission has been detected from some of the nearest spiral galaxies, where it is "normally" of the same nature as in our Galaxy, namely radiation of interstellar hydrogen and relativistic electron. The powerful radiation from the peculiar galaxies seems to be entirely due to the bremsstrahlung of electrons, which are present in high concentrations and move in strong magnetic fields. The powerful radio source Cygnus A has been identified with a galaxy 40" in diameter whose photographic apparent magnitude is $16^{m}.5$. Its radio power at 100 MHz is $5 \cdot 10^{42}$ erg/sec, which is a factor of 6 greater than its optical power. For the spiral galaxy in Andromeda this ratio is 10^{-6}. Cygnus A is regarded by some authorities as a system of two galaxies in collision. The collision between the gas clouds in the two galaxies may produce powerful radio emission, but the gas should be present in large quantities and under certain special physical conditions. Most discrete radio sources so far have not been identified with known optical objects.

B. A. VORONTSOV-VEL'YAMINOV

Chapter XIII

DUST NEBULAE

§ 80. Interstellar dust and dark nebulae. W. Struve in 1847 was the first to deduce the existence of interstellar light absorbing matter from the analysis of observational data. His great discovery, however, was not universally acknowledged by his contemporaries. At the beginning of the 20th century, Barnard suggested that the numerous dark spots that he had observed in the Milky Way were in fact dark nebulae (Figure 133 is a photograph of a dark nebula in Sagittarius), i.e., clouds of opaque matter screening off the light from distant stars. In the 1920's the distances and the sizes of numerous dark nebulae were determined, but the omnipresence of interstellar absorbing dust, even between the dark nebulae, was acknowledged only in 1930. Trumpler discovered a systematic discrepancy in the distances to open star clusters as measured from their apparent angular diameter and the apparent magnitudes of the constituent stars. This discrepancy increased with increasing distance and thus could only be attributed to increasing absorption of light along its path to the observer. This absorption was also found to be highly selective: it increased at shorter wavelengths.

Hubble's "zone of avoidance" along the galactic equator, where no galaxies are observed, analysis of absorption at various galactic latitudes, and photographs of spindle-shaped galaxies with a dust lane in their plane of symmetry (first noted by Curtiss back in 1918) led to a definite conclusion that the absorbing matter and dark nebulae markedly concentrated in the plane of the Galaxy. Soon after that the light absorption was found to be nonuniform in various directions in the galactic plane.

Fine dust is clearly the only agent which will give effective absorption despite its small mass.

The main studies of interstellar absorbing dust are concerned with the dependence of absorption on wavelength, the determination of absorption at various wavelengths in various directions and various distances from the Earth, determination of the distances, sizes and structure of the dense dark nebulae, elucidation of the spatial distribution of dust in galaxies and in the intergalactic space, analysis of the interaction of interstellar dust with diffuse gas and with stars, and the ability of dust particles to grow.

Knowledge of the absorption as a function of wavelength is highly significant, since generally absorption is measured only for a certain region of the spectrum, whereas actual observations of stars are carried out at different wavelengths. Furthermore, this dependence constitutes a virtually unique fingerprint of the absorbing particles.

FIGURE 133

According to the general law of light absorption, the intensity I_0 of light passing through an absorbing medium of optical thickness τ is reduced to

$$I = I_0 \cdot e^{-\tau}. \tag{1}$$

Let the optical thickness of the medium per unit path length be k, so that over a path of length r we have $\tau = kr$.

In the simplest case we may take

$$k(\lambda) = a + b\lambda^{-n}, \tag{2}$$

where a and b are constants; a corresponds to general (neutral) absorption, which is constant at all wavelengths. For particles much smaller than the wavelength, we have Rayleigh's law $k \sim \lambda^{-4}$.

The index n is best determined by a spectrophotometric comparison of two stars with identical true energy distribution in the spectrum which are located at greatly differing distances from the observer.

Interstellar absorption will produce a difference of Δm stellar magnitudes in the intensities of some wavelength λ in the spectra of these stars, where

$$\Delta m(\lambda) = 1.086\,[(a_1 - a_2) + (b_1 - b_2)\,\lambda^{-n}], \tag{3}$$

since by (1)

$$m - m_0 = 2.5\tau \lg e = 1.086\tau. \tag{4}$$

Comparison of the spectra of the two stars generally gives the spectrophotometric gradient $d\Delta m(\lambda)/d(1/\lambda)$, so that differentiating (3) we get

$$\frac{d\Delta m(\lambda)}{d\left(\frac{1}{\lambda}\right)} = -1.086n(b_1 - b_2)\lambda^{-(n-1)}. \quad (5)$$

If two stars of different temperatures emit as a blackbody, the spectrophotometric gradient in the visible spectrum is virtually constant and independent of wavelength. The difference in the intensities of corresponding wavelengths in the spectra of these stars therefore linearly increases with decreasing $1/\lambda$. If $n = 1$ in (5), we have the same case and the effect of selective absorption cannot be separated from the temperature effect. This naturally involves certain difficulties. Indeed, early-type supergiants are suitable for spectrophotometric comparison and they are furthermore sufficiently bright to be visible over large distances. The effect of interstellar absorption is more pronounced for these stars. However, we do not know whether or not the true energy distribution in the spectra of these supergiants coincides with that in the spectra of lower luminosity stars of the same spectral type which are closer to the Sun.

Spectrophotometry can be conveniently replaced by photoelectric measurements, which give a high accuracy in the entire spectrum from 3500 to 20,000 Å.

Between 4000 and 7000 Å, n was found to be very close to unity, so that absorption increases as $1/\lambda$. At shorter and longer wavelengths, however, the absorption function changes markedly. Beyond this spectral region the absorption remains virtually constant regardless of wavelength.

This absorption function is applicable in any direction, i.e., the physical properties of the absorbing dust are nearly identical everywhere. However, more detailed studies are needed, e.g., for stars in the Orion Nebula, where a most peculiar absorption function was found.

The parameter a in Eq. (2) cannot be estimated with any accuracy; it only seems to be fairly small, so that the general absorption of light is low (Florya). Therefore, generally one writes

$$k(\lambda) = b \cdot \lambda^{-1}. \quad (6)$$

The observations give the color excess CE as the difference between the apparent color index CI and the true color index CI_0 of the corresponding spectral type:

$$CE = CI - CI_0,$$

where the color indices are defined as the difference in the stellar magnitudes of the object at two wavelengths λ_1 and λ_2 (different systems can be used).

If the absorption in stellar magnitudes at these wavelengths is A_{λ_1} and A_{λ_2}, we have

$$CE = A_{\lambda_1} - A_{\lambda_2}.$$

By (4)

$$A = 1.086\tau = 1.086 \int_0^r k\, dr.$$

Thus, if $k(\lambda_1)/k(\lambda_2)=\mu$, we have

$$CE = 1.086\,(1-\mu)\int_0^r k(\lambda_2)\,dr,$$

or

$$CE = (1-\mu)\,A_{\lambda_2},$$

and

$$A_{\lambda_2} = \gamma\,CE, \tag{7}$$

where

$$\gamma = \frac{1}{1-\mu}. \tag{8}$$

In general

$$\gamma = \frac{\lambda^{-1}}{\lambda_1^{-1}-\lambda_2^{-1}} \tag{9}$$

is a factor which converts the color excess measured between any two wavelengths λ_1 and λ_2 to the total absorption A at wavelength λ_2 from Eq. (4).

In practice, the determination of absorption form the color excess using Eq. (7) is more accurate than direct measurement of this quantity, since it is virtually impossible to compare with any precision the distances to far objects without comparing their true and apparent magntides, which are naturally distorted by interstellar absorption.

At wavelengths far from 4000—7000 Å, one should take into consideration the true dependence of k on λ, and Eqs. (9) and (7) are no longer valid.

The most extensive series of color excess measurements of distant B-type stars (1332 stars) was obtained photoelectrically by Stebbins, Whitford, and Huffer. They measured the stellar magnitudes in the so-called C_1 system, which corresponds to wavelengths of 4260 and 4770 Å. In this system they obtained $\gamma = 7$ for visual absorption and $\gamma = 7.7$ for photographic absorption. International photographic-photovisual color excesses and photographic absorption give for γ a value somewhat less than 5. In 1950's extensive photoelectric observations of color excesses were carried out by Hiltner and others.

Note that since the interstellar absorption is virtually independent of λ at long wavelengths, the absorption does not decrease at $\lambda > 20{,}000$ Å (as long as we remain shortward of radio waves) and the Galaxy is no more transparent at these wavelengths than at $\lambda = 20{,}000$ Å.

The function $k(\lambda)$ found from observations is virtually the same for dark nebulae and the interstellar medium. This fact, combined with other considerations, shows that electrons and gas atoms which scatter light as λ^{-4} (these particles are much smaller than the light wavelengths) do not make a substantial contribution to interstellar absorption. The absorption is mainly due to minute particles of radius a comparable to the wavelength. The theory of interaction of these particles with electromagnetic waves was developed by Mie. It is fairly simple, however, only when $\alpha = 2\pi a/\lambda$ is close to unity.

Pure dielectrics with zero electrical conductivity produce true scattering. In conductors and semiconductors, electromagnetic energy is converted to heat; part of the energy is scattered, and part undergoes true absorption. If s is the density of the particles, n is the complex refractive index of the dielectric, and m is the complex refractive index of the conductor, the absorption coefficient per unit mass for a much less than unity is

$$k(a, \lambda) = \frac{32\pi^4 a^3}{s\lambda^4}\left(\frac{n^2-1}{n^2+2}\right)^2 \text{ for dielectrics}$$

and

$$k(a, \lambda) = \frac{6\pi}{s\lambda} \operatorname{Im}\left(\frac{1-m^2}{m^2+2}\right) \text{ for metals,}$$

where Im is the imaginary part of the function in parentheses.

Comparison of observations with the calculated absorbance of particles and its dependence on λ show that even though the interstellar dust particles may have a substantial spread of sizes, particles with diameters from 10^{-4} to 10^{-5} cm definitely predominate.

The variation of absorption with wavelength and the amount of absorption suggest that these are metallic particles. They are capable of producing the observed absorption at sufficiently low masses. However, the albedo of the interstellar particles obtained from observations of bright dust nebulae is too high for metals. It is close to 1, i.e., to the albedo of snow or frozen gases. Moreover, the content of metals in stars and gaseous nebulae is negligible, and it is highly improbable that the interstellar dust is of pure metallic composition.

The observed properties of the interstellar dust apparently can be satisfied by a certain mixture of dielectric particles of various diameters of the order of $\sim 0.8\mu$, which contains a certain proportion of ice particles (mainly hydrogen ice).

The minute dust particles, however, are highly sensitive to radiation pressure, especially near giants. The radiation pressure produces different effects depending on the prevalent wavelengths, the particle sizes, and particle composition (dielectric or metal).

Analysis of observations shows that the density of the absorbing interstellar medium decreases exponentially with distance from the plane of the Galaxy, so that the linear absorption coefficient becomes smaller away from the galactic plane. By analogy with the scale height of the Earth's atmosphere, we can define the equivalent halfwidth of the absorbing layer. A layer of this thickness, if its uniform density is equal to the density of the absorbing medium in the plane of the Galaxy, produces the same absorption as the real interstellar medium in the direction to the poles of the Galaxy. Parenago proposed the following formula for the calculation of absorption:

$$A(r, b) = \frac{a_0 \cdot \beta}{\sin b}\left(1 - e^{-\frac{r \sin b}{\beta}}\right).$$

Here r is the distance to the object, b is the galactic latitude of the object, and a_0 is the absorption per kiloparsec in the given direction. The values of a_0 were determined by Parenago for various directions and plotted on a map of the sky. The distance r, corrected for interstellar absorption,

is calculated from the obvious equality

$$5 \lg r = m - A(r, b) - M + 5.$$

Since r is the true distance corrected for absorption, which is not known to begin with, it is generally found by trial and error. Parenago's method uses graphs which correlate the true distance modulus $m_0 - M$ (corrected for absorption) with the apparent modulus $m - M$ for various a_0 and b. Here a_0 is taken corresponding to the effective wavelength in the relevant spectral region.

Note, however, that although the introduction of a directional quantity a_0 improves the results of calculations, it essentially contradicts the basic model of the absorption coefficient decreasing away from the galactic plane.

Thus, the mean a_0 for the galactic plane is 3.4 stellar magnitudes per kiloparsec, wheras the values of a_0 in the direction of the galactic equator are sometimes higher than this mean value.

Another extremely significant point is that Parenago's formula does not allow for the patchiness of absorption along the line of sight. Indeed the absorption coefficient in a given direction is not a monotonic function of a_0 (when reduced to the plane of the Galaxy); it is indicatory of a medium with irregular density variation. At some distance from the Sun the absorption comes out low, then for stars immersed in a large dust cloud it will rapidly increase with distance, and beyond the cloud it will remain fairly constant over large distances. Unfortunately, the absorption data are very meager even to distances of 1500—2000 parsecs, and the spatial distribution of absorption cannot be established with any certainty.

The distances and the diameters of individual sharply outlined dark nebulae are usually determined by comparing the count of stars up to a given stellar magnitude in the area of the cloud and in the adjoining transparent area. Ogorodnikov and Dobrovol'skii have shown that the luminosity function must be allowed for in these counts. We know, however, that the luminosity function is variable in different parts of the Galaxy, and reliable information on this function is available only in the neighborhood of the Sun. More accurate data on dark nebulae are therefore provided by color excess studies of stars of known luminosity. Unfortunately, O and B stars are relatively numerous among the stars which are visible over great distances, but their individual luminosities and hence distances are often highly uncertain. This greatly interferes with proper pinpointing of the dark nebulae and determination of their apparent width and line-of-sight depth. Note that near the galactic equator, where no B stars are observed (either because there are none in this direction or because they are obscured by dark clouds), the absorption cannot be found either.

Stars which begin revealing pronounced effects of absorption can be used to estimate the distance to the front edge of the dark nebula. The total absorption in the nebula is found by comparing the rapidly changing color excesses of these stars with those of the more distant stars, which are identified by slowly increasing color excess.

The nearest dark nebulae are distant about 100 pc from the Sun, and they are most prominent against the light background of the Milky Way, where they show as large black areas. These include the dark nebulae in Taurus and the "coal sacks" in the Southern Hemisphere. They often form

clusters of dark nebulae, as in Taurus. In Aquila such a cluster begins at a distance of 200 pc from the Sun and extends over more than 1000 pc along the line of sight. The separation of the Milky Way into two arms and the wide dark lane between them in the direction of the galactic nucleus are attributed to a concentration of clusters of dark nebulae in this region. The photographic absorption between these nebulae and the center of the Galaxy is over 6 stellar magnitudes.

The color excesses in some areas in Cygnus and Sagittarius reach $0^m.5$ on the C_1 scale, so that B stars often have the orange color of K stars.

The diameters of the dark nebulae range between wide limits, but the larger the nebulae the less frequent. For example, the Coal Sack is about 8 pc in diameter, numerous nebulae are of smaller size, whereas the clusters in Taurus, Auriga, and Ophiuchus are no less than 30—40 pc in diameter. Photographic absorption in these nebulae respectively varies from 1^m to 3^m.

Numerous dark nebulae, like those in Ophiuchus, extend in the form of dark long channels in the Milky Way, and quite probably we view some of the dark nebulae along the axis of such a channel.

This explains the total absence of known distant objects in certain directions and the existence of visibility corridors between them. Through one of these corridors, that in Cygnus, we see cepheids distant more than 10,000 pc which show very small reddening, this being the result of negligible absorption in this direction. However, no extragalactic nebulae are visible through this corridor. This means that farther away the absorption nevertheless markedly increases.

There are other corridors terminating with windows into the extragalactic space. The distant galaxies are clearly visible through these windows.

Ambartsumyan developed the theory of flucutation of the number of dark nebulae along the line of sight and studied the effect of these flucutations on the observed surface brightness of the Milky Way and other galaxies.

This theory applies the laws of statistics to a random distribution of objects. Earlier theories in stellar astronomy conversely assumed models with regular distribution of objects, e.g., concentric ellipsoids of varying density, and others.

A real galaxy is an object of complex structure, with spiral arms and star clouds in the gaps between them. This structure does not fit within the framework of the usual order or disorder concepts, and it requires painstaking topographic and morphological studies. Theories using different models clearly provide an indication of the deviations of the models used from the actual state of things.

Two of these models are amenable to exact mathematical treatment. One model assumes the existence of a tenuous plane-parallel layer in which the dust clouds — nebulae — float. The second model, on the other hand, starts with a random distribution of dust nebulae of various sizes with vacuum between them.

These models clearly constitute the two extremes of the actual range. Dust matter is probably reminiscent of cigarette smoke rising through air, with occasional condensations and rarefactions and traces of turbulence.

The turbulence of the interstellar dust is sustained by the differential galactic rotation, formation of new dust clouds, interaction with stars, radiation pressure, and other factors.

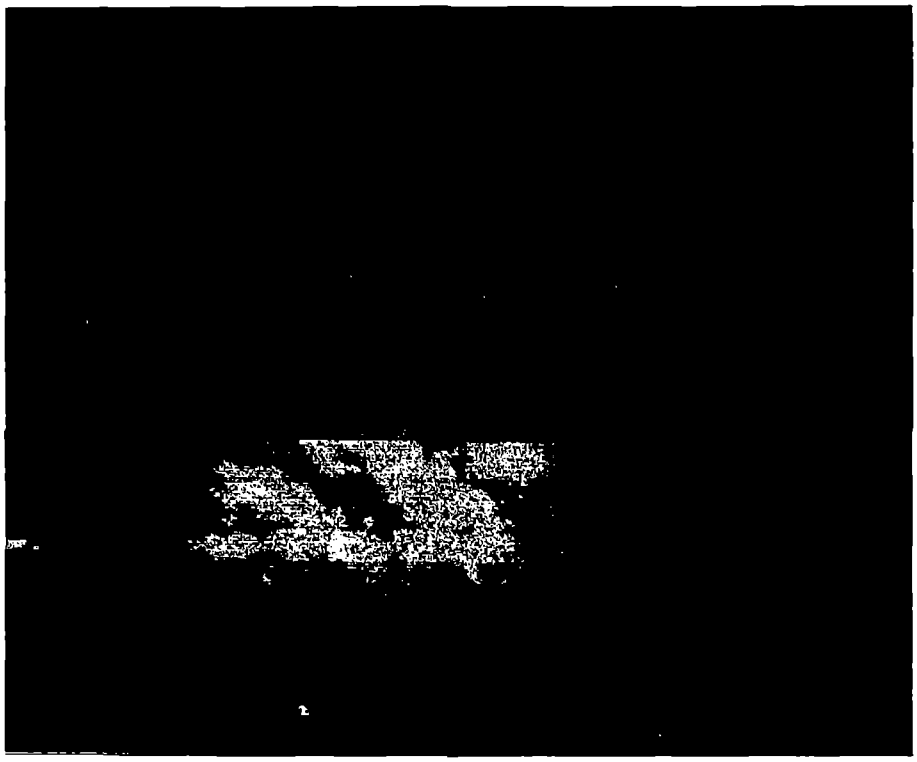


FIGURE 134

In 1947 Bok and Rayleigh discovered very small but highly absorbing dark nebulae, which were visible only against the background of some bright diffuse nebulae (the two forms are probably associated in space). These are sharply delimited nebulae of irregular shapes, although some are spherical. They were called globules (Figure 134). The smallest globules reach 10,000 a.u. in diameter, whereas the largest have diameters of 1 pc. Absorption in globules reaches 5^m, i.e., they should be extremely dense.

Combining the known absorption of dark nebulae and their diameters with Mie's theory enables us to estimate the masses of the dust clouds. For particles $10^{-4} - 10^{-5}$ cm in diameter, the minimum mass of the dust clouds is estimated by the following figures (in solar masses):

Small globules	0.002
Large globules	0.05
Coal Sack	13
A large cluster of nebulae	300

If the actual spread of the particle sizes and gas impurity are taken into considerations, higher masses are obtained. The mass of globules, apparently, is nevertheless too low to allow star formation processes.

§81. Bright dust nebulae. Ambartsumyan and Gordeladze interpreted the bright dust nebulae as the result of a chance encounter between a dark nebula and a sufficiently luminous star. If this is so, the number of bright dust nebulae illuminated by stars of a given spectral type should be proportional to the probability of finding a nebula within one of the volumes illuminated by the relevant stars. The illumination is sufficient to make a dust nebula visible only within a sphere of a certain radius centered around the star. The radius of this illumination sphere is proportional to the square root of the star's luminosity. If V is the volume of the sphere, and the number of stars of a given type in the Galaxy is n, the total illuminated volume is nV. For main-sequence stars, the luminosity function gives the number of stars of each spectral type, since the spectrum is a single-valued function of luminosity.

To compare the counts obtained in this way with observations, remember that we are only counting stars up to a certain magnitude, so that the observed volume for the high-luminosity O stars is naturally larger than the volume in which the weaker A stars are counted.

Table 1 lists the relative illumination volumes nV for stars of different spectral types and compares them with the observed number of bright diffuse nebulae illuminated by stars of the corresponding types (Ambartsumyan and Gordeladze).

TABLE 1

Spectral type	N	$nV \cdot 10^4$	Spectral type	N	$nV \cdot 10^4$
O	11	0.2	F	2	0.25
B0	7	0.6	G	1	0.18
B1—B9	54	2.9	K	2	0.25
A	5	0.8	M	0	0.02

The two columns show a good fit, if we ignore the O and B0 stars which illuminate gaseous, as well as dust, nebulae. The table therefore confirms the chance encounter hypothesis. Gaseous nebulae seem to be genetically, and not accidentally, linked with their exciting stars. The sum of the relative illuminated volumes in the table is 1/2000 of the corresponding total interstellar volume. The number of dark (unilluminated) dust nebulae should thus be a factor of 2000 higher than the number of bright (illuminated) nebulae. Hence it follows that a plane galactic layer some 200 pc thick should contain about 200,000 dark dust nebulae.

If we assume an average radius of 3 pc for these nebulae, the line of sight will intercept on the average four dark nebulae over a distance of 1 kpc in the plane of the Galaxy. This gives the mean absorption in one nebula (~ 0.3 mag) and the total absorption over any distance as the result of random fluctuations in the number of intercepted nebulae.

The above calculations should be corrected for the luminosity functions of the different spectral types, especially in view of the existence of A-type, F-type, and other supergiants, and for interstellar absorption. For example, supergiants of 5 mag are often greatly attenuated by interstellar absorption. These factors should naturally alter the figures in the table,

especially since the number of discrete nebulae is estimated fairly arbitrarily. The data on hand, however, are still insufficient to justify this improved analysis.

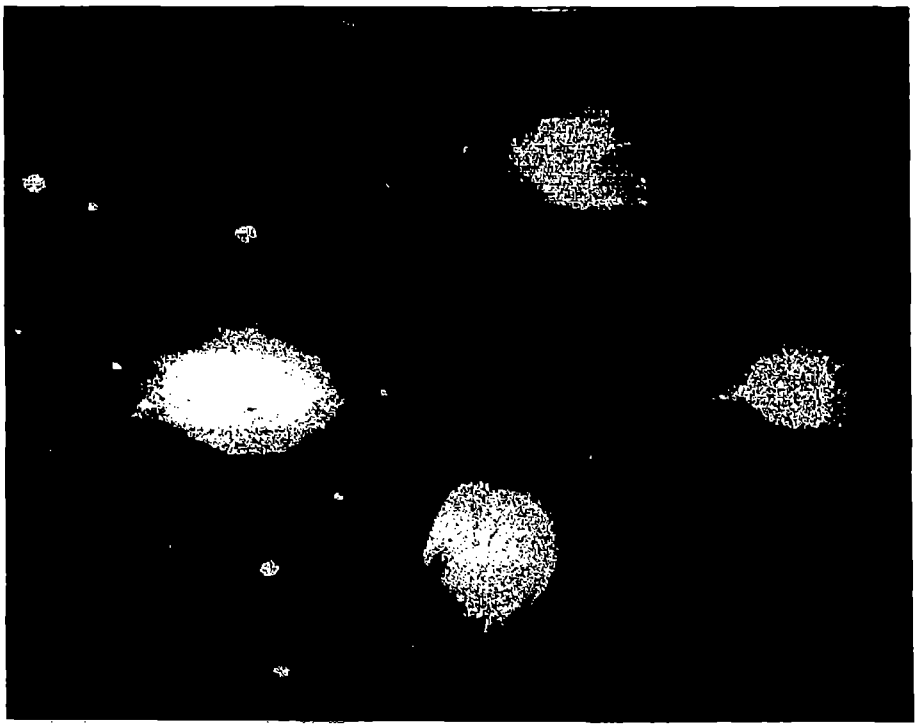


FIGURE 135

Bright nebulae (Figure 135) are no easier to study than dark nebulae, despite the fact that they are visible. One generally measures the distribution of brightness, polarization, light absorption, and compares their color with the color of the illuminating star. The results are reduced to give the particle size, the spatial position of the nebulae, and the density distribution. Numerous theoretical studies have been published, but their application to dust nebulae often does not provide the correct answer in view of the great complexity of the problem. Thus, theoretically a nebula composed of fine dust particles which lies in front of a star should appear bluer than the illuminating star if the absorption coefficient increases with decreasing wavelength. Yet the blueness of the Cygnus Nebula (color index $-0^m.5$) illuminated by a star with color index $-0^m.2$ was found to be linked up with a gaseous impurity emitting at the λ 3727 Å [OII] line. If the nebula is more distant than the star, its color should be comparable with the color of the star. For coarse particles, which are incapable of selective scattering, the phase law should be taken into consideration. Thus, particles at the periphery of the illuminated area will present a partly illuminated aspect to the observer.

The color and the light of a dust nebula depend on particle composition (whether metallic or dielectric) and on the particle size distribution, since particles of different sizes have different scattering diagrams and the albedo of dielectrics is relatively high. For particles less than 0.1 mm in diameter, most of the incident light is reflected backward, whereas for much smaller particles most of the incident flux is reflected forward.

Comparison of theory with observations indicates that high-albedo (about 0.7) particles with diameters of about 0.5μ apparently predominate.

Diffuse nebulae often show both a continuous spectrum and emission lines. In some cases, e.g., in the Orion Nebula, the continuous spectrum is definitely attributed to the dust particles. Shain, Gaze, and Pikel'ner studied about 40 diffuse nebulae on photographs taken through various light filters. They found, in addition to the previous spectrophotometric results, that it was impossible to clearly distinguish between gaseous and dust nebulae. Their calculations show that emission nebulae may contain large quantities of dust, without showing the characteristic continuous reflection spectrum. Moreover, in numerous gaseous nebulae the previously discovered continuous spectrum is associated with gas emission. For a long time the continuous spectrum of various planetary nebulae, which are definitely dust-free, remained without any satisfactory explanation. Kipper and others were apparently right in linking it up with two-photon emission from the hydrogen metastable level. In weak emission nebulae, the two-photon radiation of hydrogen and the scattered star light are apparently of comparable intensity.

Bright gaseous nebulae show dark dust spots and veinlets. The luminous mass of the nebula is often cut by dark intrusions known as "elephant trunks." These are probably the result of interaction between hot and cold dust-contaminated gas masses, which create shock waves.

Dark nebulae often have a "silver lining" along the edges. The gas may be excited to luminosity by collisions at a shock wave front, or alternatively, because of the asymmetric scattering by dust particles, a higher reflected flux reaches the observer from the edges of the nebula.

Some stars, mostly T Tau stars, are associated with small comet-like nebulae with their head toward the star. These nebulae have mixed spectra. In Ambartsumyan's opinion, the continuous spectrum is attributed neither to dust nor to two-photon emission but actually to ejection of protostellar matter from the star, which is still in the process of formation.

REFERENCES TO CHAPTERS XII AND XIII

1. ALLER, L. Gaseous Nebulae.— London. 1956.
2. AMBARTSUMYAN, V.A. et al. Teoreticheskaya astrofizika (Theoretical Astrophysics), §36.— Gostekhizdat. 1952.
3. DUFAY, J. Nébuleuses galactiques et matière interstellaire.— Paris. 1954.
4. VORONTSOV-VEL'YAMINOV, B.A. Gazovye tumannosti i novye zvezdy (Gaseous Nebulae and Novae).— Izdatel'stvo AN SSSR. 1948.
5. GURZADYAN, G.A. Planetarnye tumannosti (Planetary Nebulae).— Fizmatgiz. 1962.
6. MENZEL, D., J. BAKER, L. ALLER, G. SHORTLEY, M. HEBB, and L. GOLDBERG. Physical Processes in Gaseous Nebulae. [Russian translation. 1954.]
7. Problemy kosmicheskoi aerodinamiki (Problems of Cosmic Aerodynamics).— Moskva. 1954.

Part Five

THEORY OF STELLAR ATMOSPHERES AND GASEOUS NEBULAE

The theory of formation of stellar spectra constitutes the foundation of theoretical astrophysics. The achievements of quantum mechanics, which provided exact tools for computing the optical properties of many atoms and molecules, and the development of new methods in the theory of radiation scattering led to considerable advances in this branch of astrophysics in recent years. On the other hand, substantial observational material has been accumulated, which markedly promoted the theoretical interpretation of stellar spectra. This naturally greatly contributed to-our understanding of the physical processes in various celestial objects.

We will now proceed with a discussion of the theory of formation of the spectra of stars and gaseous nebulae and also briefly present the conclusions drawn from a comparison of the theory with observations. The theoretical aspects pertaining to the spectra of the Sun, the planets, comets, and the interstellar medium are treated in Volume III of this Course.

The present part comprises four chapters. Chapter XIV deals with stellar photospheres, i.e., the layers where the continuous spectrum of stars originate. Chapter XV is devoted to the stellar atmospheres, where the line spectrum forms. In these two chapters the treatment is based in the first approximation on the assumption of local thermodynamic equilibrium (e.g., the distribution of atoms over states is described by Boltzmann and Saha equations). This approximation greatly simplifies the calculation of stellar spectra.

Chapter XVI describes the physical processes in gaseous nebulae. The nebular spectrum comprises isolated bright lines on a weak continuous background. The nebulae emit in the visible spectrum by degradation of the high-frequency stellar radiation, and the assumption of thermodynamic equilibrium is no longer valid. However, because of the exceedingly low density of radiation and matter in nebulae, the analysis of physical processes in these systems does not present particular difficulties.

Chapter XVII is concerned with stars with bright spectral lines. These stars eject gaseous matter and therefore they are surrounded by an expanding envelope. The emission mechanism of these envelopes is fundamentally similar to the emission of nebulae, although quantitatively it is somewhat more complex.

The general physical concepts used in the following treatment were tackled in the first chapter of Vol. I of this course.

V.V. SOBOLEV

Chapter XIV

STELLAR PHOTOSPHERES

§82. Radiative equilibrium of a stellar photosphere. The photosphere is that layer of a star which emits the continuous radiation actually reaching the observer. Below the photosphere lies the stellar interior, which is inaccessible to observations, and above it extends the atmosphere which produces a line spectrum. The division of the surface layers of a star into photosphere and atmosphere is clearly artificial, but it is nevertheless highly convenient in theoretical and applied work.

The main object of the theory of stellar photospheres is the analysis of the radiation field in the photosphere. In particular, the theory should give the energy distribution in the continuous spectrum of the star. Comparison of theoretical spectra with actual observations provides certain information on the physical constitution of stellar photospheres.

The fundamental assumptions of the theory of photospheres are the following:

1. The photosphere has no energy sources and sinks. More precisely, the energy is generated in the stellar interior and is only transmitted by the photosphere on its way out. This assumption is quite reasonable, since the physical conditions in stellar photospheres are not particularly far from the conditions attainable in terrestrial laboratories, where generation of tremendous quantities of energy is definitely ruled out.

2. Energy is transferred from the stellar interior to the outside by means of radiation. Calculations show that energy transfer by heat conduction and convection is of no significance in stellar photospheres.

3. The star is in a stationary state. This assumption is not applicable only to a very small minority of stars, which reveal rapid fluctuations of magnitude and spectral type (e.g., novae). Discussion of these stars is deferred to a later stage.

A medium in which the above assumptions hold true is said to be in radiative equilibrium, and the theory based on these assumptions is called the theory of radiative equilibrium. In the following we are concerned exclusively with the theory of radiative equilibrium of stellar photospheres.

Let us write out the fundamental equations of the theory. As we have already noted, radiative transfer of energy predominates under conditions of radiative equilibirum. We should therefore use the equation of transfer (see Vol. I, Ch. I)

$$\frac{dI_\nu}{ds} = \varepsilon_\nu - \alpha_\nu I_\nu, \tag{1}$$

where I_ν is the intensity of radiation, ε_ν is the emission coefficient, α_ν is the absorption coefficient (at frequency ν). Equation (1) shows how the radiation intensity varies along the path s.

Each volume element in a state of radiative equilibrium emits precisely what it absorbs. The quantity of energy abosrbed by unit volume from all directions at all frequencies is

$$\int_0^\infty \alpha_\nu \, d\nu \int I_\nu \, d\omega,$$

and the quantity of energy emitted by this volume in all directions at all frequencies is

$$4\pi \int_0^\infty \varepsilon_\nu \, d\nu,$$

so that

$$4\pi \int_0^\infty \varepsilon_\nu \, d\nu = \int_0^\infty \alpha_\nu \, d\nu \int I_\nu \, d\omega. \tag{2}$$

Eqs. (1) and (2) are the fundamental equations of the theory of radiative equilibirum of stellar photospheres.

In the initial stages of the development of this theory, a constant absorption coefficient was assumed. We will adopt this assumption in the first stage of our discussion (putting $\alpha_\nu = \alpha$). Integrating Eq. (1) over the frequencies and putting

$$\int_0^\infty I_\nu \, d\nu = I, \qquad \int_0^\infty \varepsilon_\nu \, d\nu = \varepsilon, \tag{3}$$

we obtain

$$\frac{dI}{ds} = \varepsilon - \alpha I. \tag{4}$$

Equation (2) takes the form

$$4\pi\varepsilon = \alpha \int I \, d\omega. \tag{5}$$

The parameters α and ε entering equations (4) and (5) are functions of the distance r from the center of the star only, and the integrated radiation intensity I is a function of the distance r and the angle ϑ between the radius-vector and the beam direction. Making use of this fact, we rewrite Eq. (4) in the form

$$\cos\vartheta \frac{\partial I}{\partial r} - \frac{\sin\vartheta}{r} \frac{\partial I}{\partial \vartheta} = \varepsilon - \alpha I. \tag{6}$$

However, the equation of transfer (6) is used only for stars with extende photospheres (e.g., Wolf—Rayet stars), which will be treated separately. In the majority of "normal" stars (in particular, the Sun) the photosphere is thin compared to the star's radius. For normal stars the curvature of

the photospheric layers is thus ignorable, and they can be regarded as planeparallel. In this case the angle ϑ remains constant along the beam, and we get from Eq. (6)

$$\cos\vartheta\,\frac{\partial I}{\partial r}=\varepsilon-\alpha I. \tag{7}$$

Note that Eq. (7) is obtained directly from Eq. (4) by putting $ds=dr\sec\vartheta$ (Figure 136).

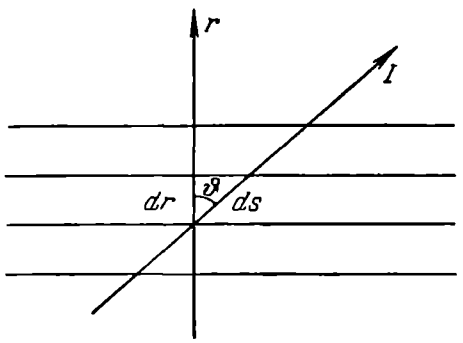


FIGURE 136

To simplify Eqs. (5) and (7), we change over from r to a new independent variable, the so-called optical depth τ:

$$\tau=\int_r^\infty \alpha\,dr. \tag{8}$$

Furthermore, we take

$$\frac{\varepsilon}{\alpha}=B. \tag{9}$$

Eqs. (5) and (7) can now be written in the form

$$\left.\begin{aligned} B&=\int I\,\frac{d\omega}{4\pi}\,,\\ \cos\vartheta\,\frac{dI}{d\tau}&=I-B. \end{aligned}\right\} \tag{10}$$

In Eqs. (10), B is a function of τ, whereas I is a function of τ and ϑ. Seeing that in spherical coordinates the solid angle element is $d\omega=\sin\vartheta\,d\vartheta\,d\varphi$, and integrating over φ from 0 to 2π, we obtain from (10)

$$\left.\begin{aligned} B(\tau)&=\frac{1}{2}\int_0^\pi I(\tau,\vartheta)\sin\vartheta\,d\vartheta,\\ \cos\vartheta\,\frac{dI(\tau,\vartheta)}{d\tau}&=I(\tau,\vartheta)-B(\tau). \end{aligned}\right\} \tag{11}$$

The theory of radiative equilibirum of stellar photospheres with frequency-independent absorption coefficient thus reduces to the solution of Eqs. (11).

Eqs. (11) and analogous systems often encountered in astrophysics served as the subject of numerous studies. Several methods have been developed for the solution of these systems, some of which are considered in the following.

First approximate method /1,2/. We introduce the mean intensity of the radiation propagating in the upward direction and the mean intensity of the radiation in the downward direction. Using $I_1(\tau)$ and $I_2(\tau)$ for these intensities, we have

$$I_1(\tau)=\int_0^{\frac{\pi}{2}} I(\tau,\ \vartheta)\sin\vartheta\,d\vartheta,\qquad I_2(\tau)=\int_{\frac{\pi}{2}}^{\pi} I(\tau,\ \vartheta)\sin\vartheta\,d\vartheta. \tag{12}$$

The first equation in (11) can now be written in the form

$$B(\tau)=\frac{1}{2}[I_1(\tau)+I_2(\tau)]. \tag{13}$$

The second equation in (11) is multiplied by $\sin\vartheta\,d\vartheta$ and integrated from 0 to $\frac{\pi}{2}$. This gives

$$\frac{d}{d\tau}\int_0^{\frac{\pi}{2}} I(\tau,\ \vartheta)\cos\vartheta\sin\vartheta\,d\vartheta=I_1(\tau)-B(\tau). \tag{14}$$

The integral in the left-hand side of this equation is approximately representable in the form

$$\int_0^{\frac{\pi}{2}} I(\tau,\ \vartheta)\cos\vartheta\sin\vartheta\,d\vartheta=\frac{1}{2}I_1(\tau), \tag{15}$$

where $1/2$ is the mean value of $\cos\vartheta$, averaged over the upper hemisphere. Eq. (14) now reduces to

$$\frac{1}{2}\frac{dI_1(\tau)}{d\tau}=I_1(\tau)-B(\tau). \tag{16}$$

Multiplying the second equation in (11) by $\sin\vartheta\,d\vartheta$ and integrating from $\frac{\pi}{2}$ to π, we similarly obtain

$$-\frac{1}{2}\frac{dI_2(\tau)}{d\tau}=I_2(\tau)-B(\tau). \tag{17}$$

Eqs. (11) are thus replaced by the approximate equations (13), (16), and (17), which can be solved without difficulty.

Adding Eqs. (16) and (17) term by term and using (13), we find

$$I_1(\tau)-I_2(\tau)=F, \tag{18}$$

where F is constant. Subtracting (17) from (16) and using (18), we find

$$I_1(\tau)+I_2(\tau)=2F\tau+C, \tag{19}$$

where C is another constant.

The physical meaning of F is obvious. The radiation flux in our approximation is

$$H = \int I \cos\vartheta \, d\omega = 2\pi \left(\tfrac{1}{2} I_1 - \tfrac{1}{2} I_2 \right) = \pi (I_1 - I_2). \tag{20}$$

Comparing (18) and (20) we get

$$H = \pi F. \tag{21}$$

Eq. (21) expresses the condition of constant radiation flux in the photosphere, which is a consequence of the absence of energy sources and sinks in the photosphere. The constant F is determined by the boundary flux values, i.e., it can be found from observations.

To find the constant C, we should remember that at the boundary ($\tau=0$) the downward (incoming) radiation flux is zero. Solving (18) and (19) for $I_2(\tau)$ and seeing that $I_2(0)=0$, we find

$$C = F.$$

Inserting this result for C in (19), we find the sum $I_1 + I_2$, which when inserted in (13) gives

$$B(\tau) = F\left(\tau + \tfrac{1}{2}\right). \tag{22}$$

If the function $B(\tau)$ is known, the second equation in (11) can be used to find the radiation intensity $I(\tau, \vartheta)$. We have thus obtained a complete solution of the problem.

Second approximate method /3/. Consider the fundamental equations in the form (10). The second equation is integrated over the solid angles. Using the first equation, we obtain

$$\frac{dH}{d\tau} = 0.$$

This again implies that the flux is constant throughout the photosphere.

Multiplying the second equation by $\cos\vartheta$ and integrating over the solid angles, we find

$$\frac{d}{d\tau} \int I \cos^2\vartheta \, d\omega = H. \tag{23}$$

The mean value of $\cos^2\vartheta$ (which is $1/3$) can be taken outside the integral, and using the first equation we get in this approximation

$$\frac{4\pi}{3} \frac{dB}{d\tau} = H. \tag{24}$$

Hence

$$B = \frac{3}{4\pi} H\tau + C, \tag{25}$$

where C is a constant.

C is obtained from the boundary condition. Clearly, at the boundary

$$B(0)=\tfrac{1}{2}I_1(0), \quad H=\pi I_1(0)$$

(see Eqs. (13) and (20)). Therefore,

$$B(0)=\frac{H}{2\pi}. \tag{26}$$

Using (26), we find from (25)

$$C=\frac{H}{2\pi}.$$

Inserting this result for C in (25) and using (21), we finally get

$$B(\tau)=F\left(\tfrac{1}{2}+\tfrac{3}{4}\tau\right). \tag{27}$$

Expression (27) for $B(\tau)$ is not greatly different from expression (22) obtained by the previous method.

Reduction to integral equation. The most natural method to find an exact solution of Eqs. (11) is the following. The second equation is solved for $I(\tau, \vartheta)$ and the result is inserted for $I(\tau, \vartheta)$ in the first equation. This gives a linear integral equation for $B(\tau)$.

The second equation in (11) yields

$$I(\tau, \vartheta)=\int_{\tau}^{\infty} B(\tau')\, e^{-(\tau'-\tau)\sec\vartheta} \sec\vartheta\, d\tau' \left(\vartheta<\frac{\pi}{2}\right), \tag{28}$$

$$I(\tau, \vartheta)=-\int_{0}^{\tau} B(\tau')\, e^{-(\tau'-\tau)\sec\vartheta} \sec\vartheta\, d\tau' \left(\vartheta>\frac{\pi}{2}\right). \tag{29}$$

Inserting this expression for $I(\tau, \vartheta)$ in the first equation in (11) and changing the order of integration, we get

$$B(\tau)=\frac{1}{2}\int_{\tau}^{\infty} B(\tau')\, d\tau' \int_{0}^{\frac{\pi}{2}} e^{-(\tau'-\tau)\sec\vartheta} \sec\vartheta \sin\vartheta\, d\vartheta - \\ -\frac{1}{2}\int_{0}^{\tau} B(\tau')\, d\tau' \int_{\frac{\pi}{2}}^{\pi} e^{-(\tau'-\tau)\sec\vartheta} \sec\vartheta \sin\vartheta\, d\vartheta.$$

Putting $\sec\vartheta=x$ in the first integral and $-\sec\vartheta=x$ in the second and seeing that $\sec\vartheta \sin\vartheta\, d\vartheta=\frac{dx}{x}$, we obtain

$$B(\tau)=\frac{1}{2}\int_{\tau}^{\infty} B(\tau')\, d\tau' \int_{1}^{\infty} e^{-(\tau'-\tau)x} \frac{dx}{x}+\frac{1}{2}\int_{0}^{\tau} B(\tau')\, d\tau' \int_{0}^{\infty} e^{-(\tau-\tau')x} \frac{dx}{x}. \tag{30}$$

Since the argument of the two exponentials can be written as $-|\tau-\tau'|x$, Eq. (30) is presented in a more compact form:

$$B(\tau)=\frac{1}{2}\int_{0}^{\infty} B(\tau')\, d\tau' \int_{1}^{\infty} e^{-|\tau-\tau'|x} \frac{dx}{x}. \tag{31}$$

The kernel of Eq. (31) is an integral exponential function defined by the equality

$$Eiz = \int_1^\infty e^{-zx} \frac{dx}{x}. \tag{32}$$

Note that Eiz has a logarithmic singularity for $z=0$, and for $z \to \infty$ it goes to zero as $\frac{e^{-z}}{z}$.

Using (32), we rewrite the integral equation for $B(\tau)$ in the final form

$$B(\tau) = \frac{1}{2} \int_0^\infty Ei|\tau - \tau'| B(\tau')\, d\tau'. \tag{33}$$

Integral equation (33) has been considered by various authors. The most comprehensive analysis is that of Hopf /4/. He obtained the exact solution of (33) in the form

$$B(\tau) = \frac{3}{4} F[\tau + q(\tau)], \tag{34}$$

where $q(\tau)$ is a function which varies slightly between

$$q(0) = \frac{1}{\sqrt{3}} = 0.58 \quad \text{and} \quad q(\infty) = 0.71.$$

Let us compare the above approximate expressions for $B(\tau)$ with the exact expression (34). Comparison of Eqs. (27) and (34) shows that both give the same result at large optical depths. At the boundary ($\tau = 0$), the approximate expression (27) gives

$$B(0) = \frac{F}{2}$$

while the exact result is

$$B(0) = \frac{\sqrt{3}}{4} F.$$

Reduction to a system of differential equations. Recently Chandrasekhar has developed a new method for the solution of various problems of radiative transfer. A detailed discussion of this method will be found in his book /5/. Here we will briefly consider Chandrasekhar's method in application to Eqs. (11).

Let $\cos\vartheta = \eta$ and rewrite Eqs. (11) in the form

$$\left.\begin{aligned} B(\tau) &= \frac{1}{2} \int_{-1}^{+1} I(\tau, \eta')\, d\eta', \\ \eta \frac{dI(\tau, \eta)}{d\tau} &= I(\tau, \eta) - B(\tau). \end{aligned}\right\} \tag{35}$$

The integral can be represented by a sum using Gauss's formula for numerical quadratures:

$$\int_{-1}^{+1} I(\tau, \eta)\, d\eta = \sum_{j=-n}^{n} a_j I(\tau, \eta_j). \tag{36}$$

Here $\eta_{-n}, \ldots, \eta_{-1}, \eta_1, \ldots, \eta_n$ ($\eta_{-j}=-\eta_j$) are the roots of the Legendre polynomial $P_{2n}(\eta)$, and a_j are weights ($a_j = a_{-j}$). Representation (36) becomes progressively more exact as we take larger and larger n.

In the n-th approximation, Eqs. (35) are replaced by a system of linear differential equations of $2n$-th order:

$$\eta_i \frac{dI_i}{d\tau} = I_i - \frac{1}{2}\sum a_j I_j \qquad (i = \pm 1, \pm 2, \ldots, \pm n). \tag{37}$$

The arbitrary constants entering the general solution of this system are determined from the following condition: the solution must be free from terms which grow exponentially with τ, $I_{-i} = 0$ for $\tau = 0$ ($i = 1, 2, \ldots, n$). The radiation flux is assumed known, so that one more constant is defined.

As an example, let us solve (37) to first approximation. For $n = 1$ we have $\eta_1 = -\eta_{-1} = \frac{1}{\sqrt{3}}$, $a_1 = a_{-1} = 1$. Eqs. (37) therefore reduce to

$$\left.\begin{aligned} \frac{1}{\sqrt{3}}\frac{dI_1}{d\tau} &= I_1 - \frac{1}{2}(I_1 + I_2), \\ -\frac{1}{\sqrt{3}}\frac{dI_{-1}}{d\tau} &= I_{-1} - \frac{1}{2}(I_1 + I_2). \end{aligned}\right\} \tag{38}$$

The solution of Eqs. (38) for the above boundary conditions gives

$$B(\tau) = \frac{3}{4} F\left(\tau + \frac{1}{\sqrt{3}}\right). \tag{39}$$

Note that this simple formula gives exact values of $B(\tau)$ not only for large τ (as (27)) but also for $\tau = 0$. This property of the solution is conserved in the higher approximations also.

§83. Brightness distribution over the stellar disk. The above results make it possible to determine the radiation intensity at any optical depth. In particular, we can calculate the intensity of outgoing radiation, i.e., $I(0, \vartheta)$. The intensity of the radiation leaving the star at an angle ϑ to the normal is clearly the luminance of the stellar disk at angular distance ϑ from the center (Figure 137).

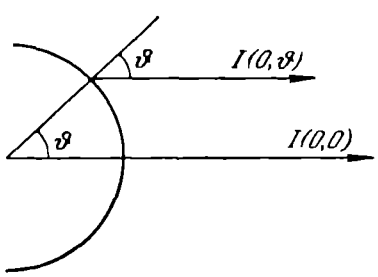


FIGURE 137

In the previous section we derived Eqs. (28) and (29) for the radiation intensity $I(\tau, \vartheta)$ in terms of $B(\tau)$. To determine the outgoing radiation intensity, we should take $\tau = 0$ in Eq. (28), which gives the intensity of upward (in our case outgoing) radiation $\vartheta < \frac{\pi}{2}$.

Writing τ for the variable of integration τ' we obtain

$$I(0, \vartheta) = \int_0^\infty B(\tau)\, e^{-\tau \sec\vartheta} \sec\vartheta \, d\tau. \tag{40}$$

Let us first determine the distribution of luminance over the stellar disk taking the approximate expressions for $B(\tau)$. Inserting first (22) and then (27) in (40), we find respectively

$$I(0, \vartheta) = F\left(\frac{1}{2} + \cos\vartheta\right) \tag{41}$$

and

$$I(0, \vartheta) = F\left(\frac{1}{2} + \frac{3}{4}\cos\vartheta\right). \tag{42}$$

It follows from Eqs. (41) and (42) that the central brightness of the disk is much greater than the limb brightness (by a factor of 3 according to (41) and 2.5 according to (42)). This result is understandable, since the radiation reaching the observer from the center of the disk originates in effectively deeper-lying photospheric layers than the limb radiation.

Let us now derive an exact expression for this limb darkening. This effect can be determined by inserting the exact expression for $B(\tau)$ in Eq. (40). The remarkable point of the entire treatment, however, is that the exact expression of limb darkening can be found even if $B(\tau)$ is not known. This possibility was first indicated by Ambartsumyan.

In his work on the theory of light scattering in planetary atmospheres, Ambartsumyan developed a method which gave the intensity of the outgoing radiation irrespective of the radiation field in the inner atmospheric layers. The essential feature of this method is the following "invariance principle": the reflectance of a medium of infinite optical thickness does not change if another layer of identical optical properties is added on top. Examination of processes which take place in an added layer of infinitesimal optical thickness leads to certain functional equations for the intensity of the outgoing radiation.

Detailed discussion of Ambartsumyan's method is unfortunately beyond the scope of the present treatment. The application of this method to our problem of brightness distribution over the stellar disk leads to the following result /6, 7/. Let $\varphi(\eta)$ be the ratio of the disk brightness at an angular distance $\vartheta = \cos^{-1}\eta$ from the center to the limb brightness; $\varphi(\eta)$ is described by the following functional equation:

$$\varphi(\eta) = 1 + \varphi(\eta)\frac{\eta}{2}\int_0^1 \frac{\varphi(\eta')}{\eta+\eta'}\,d\eta'. \tag{43}$$

Eq. (43) is readily solved by numerical methods. Some numerical results are listed in Table 1. Note that according to this table the exact ratio of the central to limb brightness is 2.91.

TABLE 1

η	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\varphi(\eta)$	1.00	1.25	1.45	1.64	1.83	2.01	2.19	2.38	2.55	2.73	2.91

Another method for calculating the intensity of outgoing radiation without considering the processes inside the transmitting medium was proposed by

the author /8/. The method is based on the derivation of linear integral equations for the intensity of reflected or transmitted radiation directly from the fundamental equations of transfer.

Multiplying the first equation in (35) by $e^{-\frac{\tau}{\eta}}\frac{d\tau}{\eta}$ and integrating from 0 to ∞, we obtain

$$I(0, \eta)=\frac{1}{2}\int_{-1}^{+1} d\eta' \int_0^{\infty} I(\tau, \eta') e^{-\frac{\tau}{\eta}}\frac{d\tau}{\eta}. \tag{44}$$

Changing over from η to η' in the second equation in (35), multiplying by $e^{-\frac{\tau}{\eta}}\frac{d\tau}{\eta}$ and integrating from 0 to ∞, we find

$$\int_0^{\infty} I(\tau, \eta') e^{-\frac{\tau}{\eta}}\frac{d\tau}{\eta}=\frac{\eta' I(0, \eta')-\eta I(0, \eta)}{\eta'-\eta}. \tag{45}$$

Insertion of (45) in (44), remembering that $I(0, \eta') = 0$ for $\eta' < 0$, gives

$$I(0, \eta)\left(1-\frac{\eta}{2}\lg\frac{1+\eta}{1-\eta}\right)=\frac{1}{2}\int_0^1 \frac{\eta'}{\eta'-\eta} I(0, \eta')\, d\eta'. \tag{46}$$

This linear integral equation with a Cauchy type kernel can be solved explicitly by standard methods.

The above theoretical law for the distribution of brightness over the stellar disk is on the whole confirmed by observations, mainly those of the Sun. Some information on limb darkening is also provided by analysis of the light curves of eclipsing variables.

Note that we have so far dealt with integrated (over all frequencies) luminances. Actual observations, however, give the distribution of luminance at certain wavelengths. A theoretical analysis of limb darkening at various wavelengths will be given in a later section.

§84. Energy distribution in the continuous spectrum. The starting assumption of radiative equilibrium in stellar atmospheres is definitely justified for all stationary stars. However, this assumption is insufficient for the development of a detailed theory of stellar photospheres, which should consider the radiation field at various frequencies. Fortunately, the actual conditions in photospheres admit of another greatly simplifying assumption, namely the hypothesis of local thermodynamic equilibrium.

As we know, thermodynamic equilibrium is established inside a constant-temperature cavity. Consider a volume element of unit optical radius at some point in the photosphere. The radiation reaching the center of this volume originates almost entirely inside the volume. Since the stellar matter is highly opaque, this volume is fairly small. The conditions inside the selected volume therefore are not greatly different from the conditions inside the radiation cavity.

The hypothesis of local thermodynamic equilibrium is clearly fully applicable to deep-lying layers of the photosphere (with optical depth of a few units). Its applicability to the upper photospheric layers naturally requires special consideration. Suitable investigations have shown that the radiation field in the photospheres of "normal" stars is indeed

consistent with the hypothesis of local thermodynamic equilibrium. As we shall see in the following, this conclusion also emerges from a comparison of the theoretical energy distribution in stellar spectra obtained from this hypothesis with actual observations.

The assumption of local thermodynamic equilibrium, however, does not apply to certain classes of "anomalous" stars. These include, e.g., Wolf—Rayet stars with extended envelopes. These stars are treated separately (see Ch. XVII).

The hypothesis of local thermodynamic equilibrium implies that any point in the photosphere is characterized by its own temperature which entirely determines the distribution of atoms over energy levels and the distribution of radiation over frequencies. In particular, in a state of local thermodynamic equilibrium, a definite relation is observed between the integrated radiation density ρ and the temperature T at any point, namely:

$$\varrho = aT^4, \tag{47}$$

where a is the Stefan constant. Since the theory of radiative equilibrium provides a relation between the radiation density ρ and the optical depth τ, Eq. (47) can be applied to determine the temperature T as a function of τ.

The radiation density is defined as

$$\varrho = \frac{1}{c}\int I\, d\omega, \tag{48}$$

where c is the velocity of light. Comparison of this equation with the first equation in (10) yields

$$\varrho = \frac{4\pi}{c} B. \tag{49}$$

Inserting for B in (47) its approximate expression from (27), we find

$$\varrho = \frac{4\pi F}{c}\left(\frac{1}{2} + \frac{3}{4}\tau\right). \tag{50}$$

Comparison of (47) with (50) yields

$$T^4 = \frac{4\pi F}{ac}\left(\frac{1}{2} + \frac{3}{4}\tau\right). \tag{51}$$

The radiation flux πF is conveniently expressed as the radiation flux from a blackbody surface at some temperature T_e. In other words, we take

$$\pi F = \sigma T_e^4, \tag{52}$$

where $\sigma = \frac{ac}{4}$. T_e is called the effective temperature of the star. It is related to the luminosity L of the star and its radius R by

$$L = 4\pi R^2 \sigma T_e^4. \tag{53}$$

Inserting (52) in (51), we obtain

$$T^4 = T_e^4\left(\frac{1}{2} + \frac{3}{4}\tau\right). \tag{54}$$

This is the final form of the expression relating the temperature T to the optical depth τ.

It follows from (54) that the surface temperature of the star is given by

$$T_0 = \frac{T_e}{\sqrt[4]{2}} = 0.841 T_e.$$

This is an approximate equality, since we have used an approximate expression for the function $B(\tau)$ in the derivation. The exact relation between T_0 and T_e is $T_0 = 0.811 T_e$.

In a state of local thermodynamic equilibrium, the Kirchhoff—Planck law should apply at any point in the photosphere:

$$\frac{\varepsilon_\nu}{\alpha_\nu} = \frac{2h\nu^3}{c^2}\frac{1}{e^{\frac{h\nu}{kT}} - 1} \tag{55}$$

Using this law and the above results of the theory of radiative equilibrium, we find the energy distribution in the continuous spectrum of a star.

Consider the equation of transfer (1). Changing over from s to a new variable r, we obtain

$$\cos\vartheta \frac{dI_\nu}{dr} = -\alpha_\nu I_\nu + \varepsilon_\nu. \tag{56}$$

Introducing the optical depth τ_ν, defined by

$$\tau_\nu = \int_r^\infty \alpha_\nu \, dr,$$

we write Eq. (56) in the form

$$\cos\vartheta \frac{dI_\nu(\tau_\nu, \vartheta)}{d\tau_\nu} = I_\nu(\tau_\nu, \vartheta) - B_\nu(\tau_\nu), \tag{57}$$

where $B_\nu = \frac{\varepsilon_\nu}{\alpha}$.

From Eq. (57) we obtain the following expression for the outgoing radiation intensity:

$$I_\nu(0, \vartheta) = \int_0^\infty B_\nu(\tau_\nu) e^{-\tau_\nu \sec\vartheta} \sec\vartheta \, d\tau_\nu. \tag{58}$$

Inserting in (58) the expression for B_ν from the Kirchhoff—Planck law, we obtain

$$I_\nu(0, \vartheta) = \frac{2h\nu^3}{c^2}\int_0^\infty \frac{e^{-\tau_\nu \sec\vartheta}}{e^{\frac{h\nu}{kT}} - 1} \sec\vartheta \, d\tau_\nu. \tag{59}$$

Eq. (59) is valid for any dependence of the absorption coefficient on frequency. First suppose, as before, that the absorption coefficient is independent of frequency (i.e., $\tau_\nu = \tau$). In this case the temperature T is expressed in terms of the optical depth τ by Eq. (54), and Eq. (59) is thus replaced by

$$I_\nu(0, \vartheta) = \frac{2h\nu^3}{c^2} \int_0^\infty \frac{e^{-\tau \sec\vartheta} \sec\vartheta \, d\tau}{e^{\frac{h\nu}{kT_e}\left(\frac{1}{2}+\frac{3}{4}\tau\right)^{-\frac{1}{4}}} - 1}. \tag{60}$$

Eq. (60), in particular, gives the brightness distribution over a stellar disk at a given frequency ν. Comparison of the theoretical brightness distribution over the disk of the Sun with actual observations at various frequencies revealed satisfactory agreement. This indicates that the assumption of frequency-independent absorption coefficient in the visible spectrum is not too improbable for the solar photosphere.

Note that Eq. (60) is often replaced by a simpler formula, which is derived as follows. Expanding $B_\nu(T)$ in powers of τ, we retain only the first two terms, i.e.,

$$B_\nu(T) = B_\nu(T_0)(1 + \beta_\nu \tau), \tag{61}$$

where

$$\beta_\nu = \frac{1}{B_\nu(T_0)} \left[\frac{dB_\nu}{dT}\frac{dT}{d\tau}\right]_{\tau=0} = \frac{3}{8}\frac{h\nu}{kT_0} \frac{1}{1 - e^{-\frac{h\nu}{kT_0}}}. \tag{62}$$

Inserting (61) in (58) we find

$$I_\nu(0, \vartheta) = B_\nu(T_0)(1 + \beta_\nu \cos\vartheta)$$

or

$$I_\nu(0, \vartheta) = I_\nu\left(0, \frac{\pi}{2}\right)(1 + \beta_\nu \cos\vartheta). \tag{63}$$

Eq. (63), however, leads to considerable errors owing to the low accuracy of the expansion (61).

Using Eq. (60), we can determine the outgoing radiation flux at frequency ν, using the standard relation

$$H_\nu(0) = 2\pi \int_0^{\frac{\pi}{2}} I_\nu(0, \vartheta) \cos\vartheta \sin\vartheta \, d\vartheta. \tag{64}$$

Inserting (60) in (64) gives

$$H_\nu(0) = \frac{2h\nu^3}{c^2} \int_0^\infty \frac{Ei_2\tau \, d\tau}{e^{\frac{h\nu}{kT_e}\left(\frac{1}{2}+\frac{3}{4}\tau\right)^{-\frac{1}{4}}} - 1}, \tag{65}$$

where

$$Ei_2\tau = \int_0^\infty e^{-\tau x} \frac{dx}{x^2}.$$

The radiation flux $H_\nu(0)$ determines the energy distribution in the continuous spectrum. It is a fundamental theoretical parameter to be compared with observation results.

Calculation of $H_\nu(0)$ from Eq. (65) leads to the following conclusion. The theoretical energy distribution in the continuous spectrum is very close to a Planckian distribution at a temperature equal to the star's effective temperature. In other words, the radiation flux $H_\nu(0)$ is approximately given by

$$H_\nu(0) = \pi \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kT_e}} - 1}. \tag{66}$$

It is only in the far ultraviolet that the radiation shows a marked excess compared to the Planckian curve, and this excess increases with increasing frequency ν.

Observations show, on the other hand, that the energy distribution in the continuous spectra of stars is definitely non-Planckian. The deviation from the Planckian distribution is not very large for the Sun and other stars of spectral types G and F, but it is highly pronounced for stars of spectral types A and B. For A and B type stars the spectrophotometric temperatures differ between wide limits for different parts of the spectrum. For example, A0 stars have spectrophotometric temperatures of about 15,000° at frequencies below the limit of the Balmer series, whereas above the Balmer limit the temperatures are about 11,000°. Another characteristic feature of A and B spectra is the jump in intensity near the Balmer limit. Thus, in the spectra of A0 stars the intensity drops to about 1/3 on crossing from frequencies below the Balmer limit to frequencies above the Balmer limit.

It is thus clear that the divergence between the theory and the observations is most considerable. Eq. (65) was derived assuming an absorption coefficient independent of frequency, and we will see in the following that without this restrictive assumption the theory does indeed show a good fit with observations.

§85. Theory of stellar photospheres with a frequency-dependent absorption coefficient. The abosrption coefficient in fact is a highly complex function of frequency, which varies for different depths in the photosphere of the same star and is of course different for different stars. Development of photosphere theory with allowance for the dependence of the absorption coefficient on frequency is therefore a difficult undertaking.

We will first consider the mathematics of this problem, regarding the absorption coefficient α_ν as a known function of frequency.

In the first approximation, our problem can be solved as follows. We start with the fundamental equations of transfer for stellar photospheres, Eqs. (1) and (2). Integrating Eq. (1) over the frequencies and writing

$$\bar{\alpha} = \frac{\int \alpha_\nu I_\nu \, d\nu}{\int I_\nu \, d\nu}, \tag{67}$$

we obtain

$$\frac{dI}{ds} = -\bar{\alpha} I + \varepsilon. \tag{68}$$

Using (67), we rewrite Eq. (2) in the form

$$4\pi\varepsilon = \bar{\alpha} \int I \, d\omega. \tag{69}$$

The $\bar{\alpha}$ defined by (67) is the absorption coefficient averaged over the frequencies with weight I_ν. Strictly speaking, the weight is an unknown function of frequency, which furthermore depends on position and direction. Therefore, in calculating the average coefficient $\bar{\alpha}$, we assume that the radiation intensity I_ν is approximately described by Planck's equation with a temperature corresponding to the given position in the photosphere. In this case $\bar{\alpha}$ is only a function of position in the photosphere.

In this approximation Eqs. (68) and (69) are no different from Eqs. (4) and (5) obtained assuming a frequency-independent abosrption coefficient. We may therefore apply all the previous results obtained by solving Eqs. (4) and (5). In particular, we may use Eq. (27) to describe the relation of temperature T to optical depth τ. In the case under consideration, the optical depth is defined by

$$\tau = \int_r^\infty \bar{\alpha} \, dr. \tag{70}$$

Calculation of energy distribution in stellar spectra is obviously of the greatest interest. These calculations require Eq. (59), which defines the function $I_\nu(0, \vartheta)$ for an absorption coefficient which is an arbitrary function of frequency. Since in this equation the temperature is a function of τ, while τ_ν is the variable of integration, we should first establish a relation between τ and τ_ν. It is difficult to establish an exact relation between the two quantities, since $\frac{\alpha_\nu}{\bar{\alpha}}$ is variable with depth. Approximately, however, we may take

$$\tau_\nu = \int_r^\infty \alpha_\nu \, dr = \frac{\alpha_\nu}{\bar{\alpha}} \int_r^\infty \bar{\alpha} \, dr = \frac{\alpha_\nu}{\bar{\alpha}} \tau, \tag{71}$$

where for $\frac{\alpha_\nu}{\bar{\alpha}}$ we take its value in the topmost surface layers of the star.

Using (71) and (27), we obtain from (59)

$$I_\nu(0, \vartheta) = \frac{2h\nu^3}{c^2} \int_0^\infty \frac{e^{-\frac{\alpha_\nu}{\bar{\alpha}} \tau \sec\vartheta}}{e^{\frac{h\nu}{kT_0}\left(\frac{1}{2} + \frac{3}{4}\tau\right)^{-\frac{1}{4}}} - 1} \sec\vartheta \, \frac{\alpha_\nu}{\bar{\alpha}} \, d\tau. \tag{72}$$

This is a fairly crude method for the determination of the radiation field in the photosphere. It nevertheless gives satisfactory results in cases when the absorption coefficient is not particularly sensitive to frequency. As we shall see in the following, this is indeed the case in the Sun's photosphere.

A rigorous theory of photospheres should proceed from simultaneous solution of Eqs. (1), (2), and (55). Changing over in Eq. (1) from s to a new

variable r, we write these equations in the form

$$\left.\begin{aligned} \cos\vartheta\frac{dI_\nu}{dr} &= -\alpha_\nu I_\nu + \varepsilon_\nu, \\ 4\pi\int_0^\infty \varepsilon_\nu\, d\nu &= \int_0^\infty \alpha_\nu\, d\nu \int I_\nu\, d\omega, \\ \frac{\varepsilon_\nu}{\alpha_\nu} &= \frac{2h\nu^3}{c^2}\frac{1}{e^{\frac{h\nu}{kT}}-1}. \end{aligned}\right\} \tag{73}$$

The absorption coefficient α_ν entering the above equations is a function of temperature and density. Since the distribution of density in the photosphere is related to the temperature distribution and is derived from the conditions of mechanical equilibrium, the equation of mechanical equilibrium must in general be added to the three equations in (73).

The absorption coefficient is often written as a product of two functions, of which one depends only on the frequency ν and the temperature T, and the other on the temperature T and the density ρ, i.e.,

$$\alpha_\nu = \Phi(\nu, T)\,\Psi(T, \varrho). \tag{74}$$

In this way, we avoid the determination of density distribution in the photosphere, since r can be replaced by a new independent variable

$$\zeta = \int_r^\infty \Psi(T, \varrho)\, dr. \tag{75}$$

The three equations in (73) are now readily reduced to a single integral equation which expresses the temperature T as a function of ζ. The constant figuring in the solution of this homogeneous equation is determined from the given total radiation flux H.

The corresponding equation is highly complex. Mustel' /9/ therefore proposed other (approximate) methods for solving Eqs. (73) subject to condition (74).

Let us now consider the dependence of the absorption coefficient on frequency. Since hydrogen is the most abundant element in the surface layers of stars, it apparently makes the main contribution to radiation absorption. We are therefore concerned primarily with absorption of radiation by hydrogen atoms.

The volume absorption coefficient of hydrogen atoms is expressed by (Vol. I, Ch. I, §9)

$$\alpha_\nu = n_e n^+ \frac{2^4\pi^2 e^6}{3\sqrt{3}ch^4}\frac{kT}{(2\pi mkT)^{\frac{3}{2}}}\left[1 + 2\frac{\chi_1}{kT}\sum_i \frac{e^{\frac{\chi_i}{kT}}}{i^3}\right]\frac{1}{\nu^3}\left(1 - e^{-\frac{h\nu}{kT}}\right). \tag{76}$$

Here n_e and n^+ is the number of free electrons and protons in 1 cm^3, respectively, χ_i is the energy of ionization from the i-th state. The first term in brackets accounts for free-free transitions; the second term for bound-free transitions (i.e., photoionization). The summation over i is performed remembering that ionization from the i-th state involves absorption of quanta with energies $h\nu > \chi_i$. Therefore, in the spectral region

beyond the limit of the Lyman series the summation begins with $i = 1$, between the Balmer limit and the Lyman limit the summation begins with $i = 2$, etc. The derivation of (76) implies validity of Boltzmann and Saha equations for the distribution of atoms over states, i.e., the assumption of thermodynamic equilibrium is utilized.

It is interesting to establish the variation of the absorption coefficient with temperature. Note that the temperature enters Eq. (76) implicitly, through n^*. Eq. (76) should therefore be considered in conjunction with the ionization equation

$$\frac{n_e n^*}{n_1} = \frac{(2\pi m kT)^{\frac{3}{2}}}{h^3} e^{-\frac{\chi_1}{kT}}. \tag{77}$$

which gives the relative number of ions as a function of T.

If the temperature in the photosphere is less than 7,000—8,000°, the great majority of hydrogen atoms are neutral. Therefore, to determine the explicit dependence of α_ν on T, we should pass in (76) from $n_e n^*$ to the number of neutral atoms n_1 (which is virtually equal to the total number of hydrogen atoms). In this case we obtain

$$\alpha_\nu = n_1 \frac{2^4 \pi^2 e^6 h^2}{3\sqrt{3}c} kT \left[e^{-\frac{\chi_1}{kT}} + 2\frac{\chi_1}{kT} \sum_i \frac{e^{-\frac{\chi_1 - \chi_i}{kT}}}{i^3} \right] \frac{1}{\nu^3} \left(1 - e^{-\frac{h\nu}{kT}}\right). \tag{78}$$

Since for hydrogen $\frac{\chi_1}{k} = 157{,}200$ and $\chi_i = \frac{\chi_1}{i^2}$, Eq. (78) shows that in the photospheres of late-type stars hydrogen atoms absorb radiation only in the Lyman continuum. This is clearly due to the negligible number of excited hydrogen atoms at low temperatures. On the other hand, the frequency distribution of radiation at low temperatures is peaked in the infrared. We thus conclude that in the photospheres of late-type stars the absorption of radiation by hydrogen atoms is negligible.

As the temperature increases, the hydrogen atoms absorb progressively more. This is associated with the increase in the number of excited atoms and with the displacement of the peak of the spectral curve toward higher frequencies. At temperatures over 8,000°, α_ν should be calculated from Eq. (76) taking n^* and n_e equal to half the number of hydrogen atoms in 1 cm^3. We see from Eq. (76) that as the temperature is further increased, α_ν slowly decreases.

As the abundance of hydrogen atoms is greater than that of all the other elements combined, the absorption of radiation in the photospheres of hot stars is probably due mainly to hydrogen. Detailed analysis has shown that this is actually the case in photospheres of stars with effective temperatures in the neighborhood of 10,000—20,000°. In the photospheres of hotter stars the absorption by helium atoms is no longer negligible.

The theory of stellar photospheres with effective temperatures of 10,000—20,000° was developed by Mustel' /9, 10/. Since the expressions for the absorption coefficient of hydrogen atoms satisfy condition (74), the calculations proved to be of reasonable complexity.

Figure 138 gives the theoretical curve plotting the energy distribution in the spectrum of a B2 star ($T_e = 20{,}000°$). The Planckian curve corresponding to the same T_e is also shown: the areas under the two curves are

equal, both being $\frac{\sigma}{\pi} T_e^4$. We see that the actual energy distribution curve in the stellar spectrum has nothing in common with the Planckian curve.

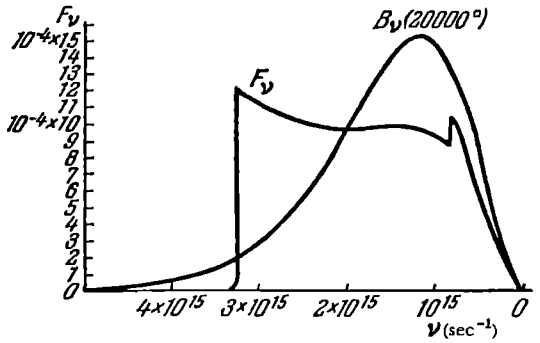


FIGURE 138

A remarkable feature are the jumps near the limits of the different series. This is a consequence of the discontinuous variation of the absorption coefficient. On crossing a series limit from lower to higher frequencies, the absorption coefficient increases discontinuously (in Eq. (76) this corresponds to the appearance of an additional term in the sum over i). By Eq. (59), the outgoing radiation flux therefore decreases discontinuously at these frequencies.

Mustel' calculated the energy distribution in the stellar spectrum and found the Balmer jump and the spectrophotometric temperatures below and above the Balmer limit. The jump is defined by the equality

$$D = \lg \frac{H_{\nu<\nu_2}}{H_{\nu>\nu_2}}, \tag{79}$$

where ν_2 is the frequency of the Balmer limit. The spectrophotometric temperature T_s is obtained from the equation

$$\frac{d}{d\nu} \lg H_\nu = \frac{3}{\nu} - \frac{h}{kT_s} \frac{1}{1-e^{-\frac{h\nu}{kT_s}}}. \tag{80}$$

The T_s values below ($\nu < \nu_2$) and above the Balmer limit will be denoted by T_s' and T_s'', respectively.

The theoretical values of D, T_s', T_s'' for stars of spectral types A0, B5, and B2 are listed in Table 2. The observed values are also given for comparison. The theory clearly shows good fit with observations. This primarily indicates that hydrogen is indeed the main absorbing agent in the photospheres of these stars. The next conclusion is that the stellar photospheres do not show great departure from local thermodynamic equilibrium.

TABLE 2

Spectral type	A0	B5	B2
T_e	10 500°	15 000°	20 000°
T'_s observed	16 500°	23 000°	26 500°
T'_s theor.	19 000	21 000	23 000
T''_s observed	11 000	16 000	19 000
T''_s theor.	10 500	15 000	19 000
D observed	0.47	0.24	0.11
D theor.	0.49	0 22	0.10

On passing from the hot A and B stars to colder stars, the absorption by hydrogen atoms diminishes. At low temperatures, radiation is mainly absorbed by atoms with low excitation and ionization potentials. The most significant of these is the negative hydrogen ion, i.e., an atom of hydrogen with one extra electron attached (H^-).

The number of negative hydrogen ions in a stellar photosphere is determined by the ionization equation with neutral hydrogen atoms regarded as the ionized particles, i.e.,

$$n_e \frac{n_1}{n^-} = \frac{(2\pi m kT)^{\frac{3}{2}}}{h^3} e^{-\frac{\chi_1}{kT}}, \tag{81}$$

where n_1 and n^- is the number of neutral hydrogen atoms and the number of negative hydrogen ions in 1 cm^3. Since the ionization energy of the negative hydrogen ion is very low ($\chi_1 = 0.75$ eV), n^- is insensitive to temperature. However, n_e and n_1 are highly sensitive to temperature, and since n^- is proportional to these concentrations, it is also affected by temperature. The photospheres of hot stars are rich in free electrons, but the concentration of neutral hydrogen atoms is very low. For this reason they contain very few negative hydrogen ions. The photospheres of very cold stars of spectral types M, N, and S are rich in neutral hydrogen atoms, but contain relatively few free electrons. Therefore, they also contain few negative hydrogen ions. It is only in the photospheres of stars of spectral types F, G, K that the number of negative hydrogen ions is high. Calculations show that in these cases absorption of radiation by negative hydrogen ions plays an important, and in G-type stars even a predominant, role.

The negative hydrogen ion has one discrete energy level, which as we have noted above is very shallow. Therefore even a quantum of very low frequency ($h\nu > \chi_1$) absorbed by H^- suffices to detach the extra electron of this ion. The frequency dependece of the H^- absorption coefficient is most peculiar. As the frequency increases, the absorption coefficient does not decrease (as, say, for atomic hydrogen): at first it actually increases (reaching a maximum in the red region) and only then starts decreasing. The absorption coefficient of H^- is plotted as a function of wavelength in Figure 139. It has been calculated for one negative hydrogen ion.

The volume absorption coefficient is derived from the equality $\alpha_\nu = n k_\nu$. Strictly speaking, a term associated with free-free transitions should be added to this expression, but in all the relevant cases it is generally small.

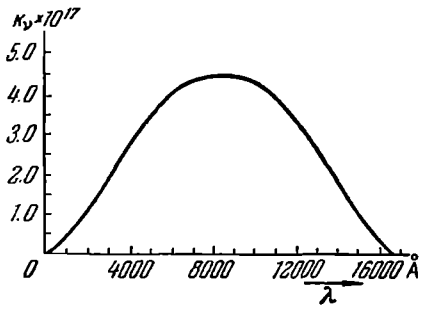


FIGURE 139

The theory of photospheres where radiation is mainly absorbed by the negative hydrogen ion was developed by Chandrasekhar and Münch /11/. They used a method not unlike that discussed at the beginning of this section. A good fit between the theoretical and the observed energy distribution was obtained for the solar spectrum. We have already noted that a satisfactory fit is ensured even by the theory assuming a frequency-independent absorption coefficient. This is so because the absorption coefficient due to the negative hydrogen ion is not particularly sensitive to frequency in the visible spectrum.

The good fit between the theoretical and the actual energy distribution in the continuous spectrum of the Sun (and other stars of this type) also supports the hypothesis of local thermodynamic equilibrium for stellar photospheres. This hypothesis has to be abandoned apparently only when considering the fine features of the spectra.

In conclusion let us consider the absorption of radiation in the photospheres of stars of other spectral types.

The absorption of radiation in the photospheres of stars colder than G-type is associated with neutral atoms and molecules of metals, as well as negative hydrogen ions. However, the theoretical energy distribution in the spectra of these stars has not been studied. Moreover, the observations do not give reliable energy distribution curves for these spectra owing to the presence of high-intensity molecular bands.

As we have noted in the preceding, the absorption in stellar photospheres with temperatures exceeding 20,000° is associated with neutral and ionized helium atoms, as well as hydrogen atoms. The expression for the absorption coefficient due to helium ions is obtained from the general expression for the absorption coefficient of hydrogen-like atoms taking $Z = 2$ (Eq. (76) corresponds to the case $Z = 1$). The expression for the absorption coefficient of neutral helium is also known. However, the complete theory of photospheres which takes simultaneous account of the absorption of radiation by hydrogen, helium, and ionized helium has not been developed so far. One particular case was treated by Underhill /12/.

In the surface layers of the hot supergiants, radiative transfer is largely governed by the scattering of radiation off the free electrons. The volume coefficient of scattering by free electrons is

$$\sigma_\nu = n_e \frac{8\pi}{3} \left(\frac{e}{mc^2} \right)^2 . \qquad (82)$$

Comparison of Eqs. (76) and (82) shows that the ratio of the free-electron scattering coefficient to hydrogen absorption coefficient is inversely

proportional to n^*, i.e., it increases as the density of photospheric gas decreases. The same relations show that the ratio $\frac{\sigma_\nu}{\alpha_\nu}$ increases with increasing temperature. In the photospheres of hot supergiants, where the density is low and the temperature is high, the ratio $\frac{\sigma_\nu}{\alpha_\nu}$ may reach values comparable with unity.

Introduction of electron scattering should clearly reduce the calculated intensity jumps at the limits of the various series, and thus alter the spectrophotometric temperatures. The observations indeed seem to reveal an effect of this kind, but the corresponding theory has hardly been developed. Some aspects of the theory of photospheres with electron scattering in application to Wolf—Rayet stars were considered by Ambartsumyan /13/ and Slyusarev /14/.

§86. The structure of stellar photospheres. One of the fundamental problems in the theory of stellar photospheres is the variation of temperature and density with depth. To solve this problem, the equations of transfer (73) should be considered in conjunction with the equation of mechanical equilibrium.

Each volume element in the photosphere is balanced by the action of three forces: 1) gravitation, 2) gas pressure, 3) radiation pressure. Consider a volume element of 1 cm^2 cross section and thickness dr. The weight of this volume is $g\rho dr$, where g is the gravitational acceleration. Let p be the gas pressure, p' the radiation pressure, $d(p+p')$ the change in pressure across the volume element on passing from r to $r+dr$. The outward pressure force is then $-d(p+p')$. Equating the weight to the force of pressure, we get

$$d(p+p') = -g\varrho\, dr. \tag{83}$$

The gas pressure p can be expressed in terms of the density ρ and the temperature T using the equation of state of ideal gases:

$$p = \frac{R}{\mu}\varrho T, \tag{84}$$

where $R = \frac{k}{m_n}$ is the gas constant, and μ is the average molecular weight. The radiation pressure in case of local thermodynamic equilibrium is given by

$$p' = \frac{1}{3}aT^4. \tag{85}$$

The equation of mechanical equilibrium (83), combined with Eqs. (84) and (85), supplements the equations of transfer of the previous section, Eqs. (73). We have thus obtained a set of equations whose solution completely describes the distribution of matter and radiation in the stellar photosphere.

The absorption coefficient α_ν entering these equations is regarded as a known function of frequency, temperature, and density. For an arbitrary function of this kind, the above system of equations is very difficult to

solve. However, if the absorption coefficient is representable in the form (74), the equations are substantially simplified. As we have noted above, Eqs. (73) alone give the temperature as a function of ζ, defined by (75). Hence we obtain the relation

$$\frac{d\zeta}{dT}\frac{dT}{dr} = -\Psi(T, \rho), \tag{86}$$

where $\frac{d\zeta}{dT}$ is known. Eq. (86) taken in conjunction with the three preceding equations constitute a system of ordinary differential equations from which the density ρ and the temperature T can be found as a function of r.

We have observed before that for atomic hydrogen α_ν satisfies condition (74). This enabled Mustel' to calculate without much difficulty the radiation field in the photospheres of stars of spectral types A0—B2. Using his results for the temperature gradient $\frac{dT}{d\xi}$, Berdichevskaya /15/ found ρ and T as a function of r for an A0 star (T_e = 10,500°). This was achieved by numerical solution of Eqs. (83—86).

A more approximate, though much more general method of investigating the structure of photospheres is available. In this method, Eqs. (73) are replaced by the approximate equations (68) and (69), where the average absorption coefficient $\bar{\alpha}$ is defined by (67) (with Planckian I_ν).

Before this method can be applied, however, we should express the average absorption coefficient in terms of ρ and T. Let us derive the corresponding equation for a hydrogen atom.

Using Eq. (76) for α_ν, we obtain

$$\int_0^\infty \alpha_\nu I_\nu \, d\nu = n_e n^* \frac{2^4\pi^2 e^6}{3\sqrt{3}ch} \frac{kT}{(2\pi mkT)^{3/2}} \frac{2h}{c^2} \int_0^\infty \left[1 + 2\frac{\chi_1}{kT}\sum_i \frac{e^{\frac{\chi_i}{kT}}}{i^3}\right] e^{-\frac{h\nu}{kT}} d\nu,$$

or changing the order of summation and integration,

$$\int_0^\infty \alpha_\nu I_\nu \, d\nu = n_e n^* \frac{2^4\pi^2 e^6}{3\sqrt{3}ch} \frac{kT}{(2\pi mkT)^{3/2}} \frac{2h}{c^2} \left[\frac{kT}{h} + 2\frac{\chi_1}{kT}\sum_{i=1}^{\infty}\frac{1}{i^3}\int_{\frac{\chi_i}{h}}^{\infty} e^{\frac{\chi_i - h\nu}{kT}} d\nu\right]$$

Integrating, we obtain

$$\int_0^\infty \alpha_\nu I_\nu \, d\nu = n_e n^* \frac{2^4\pi^2 e^6}{3\sqrt{3}ch} \frac{kT}{(2\pi mkT)^{3/2}} \frac{2h}{c^2} \frac{kT}{h}\left[1 + 2.4\frac{\chi_1}{kT}\right]. \tag{87}$$

Now,

$$\int_0^\infty I_\nu \, d\nu = \frac{2h}{c^2}\left(\frac{kT}{h}\right)^4 \int_0^\infty \frac{x^3\,dx}{e^x - 1} = \frac{2h}{c^2}\left(\frac{kT}{h}\right)^4 \frac{\pi^4}{15}. \tag{88}$$

Inserting (87) and (88) in Eq. (67) and using the relation

$$\chi_1 = \frac{2\pi^2 me^4}{h^2}$$

we obtain

$$\bar{\alpha} = \frac{40}{\pi^4 \sqrt{3}} \frac{e^2 h^4}{mc\,(2\pi m)^{3/2}} \chi_1 \left[1 + 2.4 \frac{\chi_1}{kT} \right] \frac{n_e n^+}{(kT)^{7/2}} . \tag{89}$$

Eq. (89) is valid for any hydrogen-like ion and, to some approximation, for other atoms as well.

Note that the first term in square brackets in Eq. (89) corresponds to free-free transitions, and the second term to free-bound transitions. If the radiation is mainly absorbed by hydrogen atoms, the first term predominates at temperatures above 400,000° and the second term below 400,000° (since for hydrogen $\frac{\chi_1}{k} = 157{,}200$).

If the hydrogen is completely ionized (so that $n_e = n^+ \sim \rho$), we obtain from Eq. (89) in the two extreme cases:

$$\bar{\alpha} \sim \frac{\varrho^2}{T^{7/2}} \tag{90}$$

(for reltively high temperatures) and

$$\bar{\alpha} \sim \frac{\varrho^2}{T^{9/2}} \tag{91}$$

(for relatively low temperatures). Eqs. (90) and (91) are very often used in astrophysics.

When the dependence of the average absorption coefficient on ρ and T is known, the distribution of density and temperature in the photosphere can be determined by considering the first three equations of this section together with the equation

$$dT^4 = \frac{3}{4} T_e^4 \, d\tau = -\frac{3}{4} T_e^4 \bar{\alpha} \, dr, \tag{92}$$

which follows from (54) and (70).

In the following we shall see that for the great majority of stars radiation pressure is negligible compared to gas pressure. In this case, from (83) and (84) we get

$$\frac{R}{\mu} d(\varrho T) = -g\varrho \, dr. \tag{93}$$

The problem is thus reduced to simultaneous solution of Eqs. (92) and (93).

Writing $\bar{\alpha} = \varkappa\varrho$, we will assume at this stage that $\varkappa = \text{const}$. From Eqs. (92) and (93) we then have

$$\frac{R}{\mu} d(\varrho T) = \frac{4}{3} \frac{g}{\varkappa} \frac{dT^4}{T_e^4} , \tag{94}$$

or, integrating,

$$\varrho = \frac{4}{3} \frac{g\mu}{\varkappa R} \frac{T^4 - T_0^4}{T_e^4 T} . \tag{95}$$

In the deep-lying layers of the photosphere, where $T^4 \gg T_0^4$, the density s related to temperature by

$$\varrho = \frac{4}{3} \frac{g\mu}{\varkappa R T_e^4} T^3. \tag{96}$$

Inserting (96) in (92), we obtain for the temperature gradient

$$\frac{dT}{dr} = -\frac{g\mu}{4R}. \tag{97}$$

Having used $\varkappa = \text{const}$, we obtained $\varrho \sim T^3$. Adopting this relation between ϱ and T, we find from Eqs. (90) and (91) that $\varkappa$ is not highly variable with depth. The assumption of constant $\varkappa$ was thus not very far from truth.

Eqs. (92) and (93), however, can also be solved without much difficulty under more general assumptions concerning $\bar{\alpha}$. Taking, say, $\bar{\alpha} \sim \varrho^2 T^{-s}$, where s is a parameter, we obtain

$$\varrho \sim T^{\frac{s+3}{2}} \tag{98}$$

and

$$\frac{dT}{dr} = -\frac{2}{s+5}\frac{g\mu}{R}. \tag{99}$$

Let us apply the above equations to the Sun's photosphere. Putting in (97) $g = 2.7 \cdot 10^4$, $\mu = 1$, $R = 8.3 \cdot 10^7$, we obtain $\frac{dT}{dr} = -10^{-4}$ deg/cm. In other words, the temperature increases by 10 deg as we descend 1 km into the solar photosphere. Note that the temperature gradient is independent of the absorption coefficient.

From Eqs. (92) and (97) we can derive the geometrical thickness of the layer of unit optical thickness, $\left|\frac{dr}{d\tau}\right|$. Combining the two equations, we get

$$\frac{dr}{d\tau} = -\frac{3}{4}\frac{T_e^4}{T^3}\frac{R}{g\mu}. \tag{100}$$

Let $T = T_e$. Then $\left|\frac{dr}{d\tau}\right|$ is a characteristic of the photosphere thickness. For the Sun, the photosphere is about 100 km thick. Since the radius of the Sun is 700,000 km, we see that the thickness of the photosphere is negligible compared to the radius. We have already made use of this fact at the beginning, when we treated the photospheric layers as plane-parallel.

In deriving Eq. (93) we ignored radiation pressure compared to gas pressure. Let us examine the validity of this approximation.

For the radiation pressure gradient we use the general expression

$$\frac{dp'}{dr} = -\frac{1}{c}\int \alpha_\nu H_\nu \, d\nu. \tag{101}$$

Introducing the average absorption coefficient, we get

$$\frac{dp'}{dr} = -\frac{\bar{\alpha} H}{c}. \tag{102}$$

Let us calculate the ratio of radiation pressure to total pressure. To this end we divide (102) by (83). Substituting $\varkappa\varrho$ for $\bar{\alpha}$, we get

$$\frac{dp'}{d(p+p')} = \frac{\varkappa H}{gc}. \tag{103}$$

If $\varkappa = \text{const}$, this ratio is constant throughout the photosphere.

Eq. (103) makes it possible to estimate the contribution from radiation pressure to the overall pressure. For most stars, the fractional contribution is much less than unity; for Sun-type stars it is of the order of 1/100. It is only in the photospheres of the hot supergiants that the radiation pressure is comparable with gas pressure.

REFERENCES TO CHAPTER XIV

1. SCHUSTER, A. Radiation through a foggy atmosphere.— ApJ, 21:1. 1905.
2. SCHWARZSCHILD, K. Über das Gleichgewicht der Sonnenatmosphäre.— Göttinger Nachr. 41. 1906.
3. EDDINGTON, A.S. The Internal Constitution of the Stars. 1930.
4. HOPF, E. Mathematical Problems of Radiative Equilibrium.— Cambridge Tracts in Mathematics and Mathematical Physics. Cambridge, No. 31. 1934.
5. CHANDRASEKHAR, S. Radiative Transfer. 1950.
6. AMBARTSUMYAN, V.A. Scattering of light by planetary atmospheres.— Astron. Zh., 19:30. 1942.
7. AMBARTSUMYAN, V.A. Diffusion of light through a medium of high optical thickness.— Doklady AN SSSR, 43, No.3. 1944.
8. SOBOLEV, V.V. Brightness distribution over the stellar disk.— Astron. Zh., 26, No.1. 1949.
9. MUSTEL', E.R. Theory of radiative equilibrium of stellar atmospheres with a frequency-dependent absorption coefficient.— Trudy, Shternberg Astron. Inst., 12, No.2. 1940.
10. MUSTEL', E.R. Energy distribution in the continuous spectrum of early-type stars.— Astron. Zh., 18:297. 1941; 21:133. 1944.
11. CHANDRASEKHAR, S. and G. MUNCH. The continuous spectrum of the Sun and the stars.— ApJ, 104:446. 1946.
12. UNDERHILL, A. On Model Atmosphere for the High-Temperature Stars.— Copenhagen. 1950.
13. AMBARTSUMYAN, V.A. Thompson scattering in stellar atmospheres.— Uchenye Zapiski, Leningrad University, No.22. 1938.
14. SLYUSAREV, S.G. Scattering of light by free electrons in stellar photospheres.— Doklady AN SSSR, 95, No.4. 1954.
15. BERDICHEVSKAYA, V.S. Structure of photospheres of A0-type stars.— Astron. Zh., 22:359. 1945.

V. V. SOBOLEV

Chapter XV

STELLAR ATMOSPHERES

§87. The Schwarzschild—Schuster model. Initially, we will regard the outermost layers of a star as divided into two components, the photosphere and the atmosphere. The photosphere limits continuous radiation, without any absorption lines, and it is on passing through the atmosphere that the stellar spectrum develops absorption lines.

This basic model of the stellar envelope is known as the Schwarzschild—Schuster model. It is applicable since the absorption coefficient in a spectral line is much greater than the absorption coefficient in the continuous spectrum. Therefore, in the outermost strata of a star, absorption and emission in the continuous spectrum are of no great significance, whereas absorption and emission in spectral lines are still quite strong.

Let us consider the formation of an absorption line corresponding to the transition from an i-th to a j-th state in a given atom. Let k_ν be the absorption coefficient of a quantum of frequency ν inside the line, calculated from one atom (for simplicity we omit the subscripts i and j, although they are always implied). Further, let $n(r)$ be the number of i-th state atoms in 1 cm^3 at a distance r from the center of the star (again for simplicity the subscript i is omitted). The optical thickness of the atmosphere at the frequency ν is given by

$$\tau_\nu^0 = \int_{r_0}^{\infty} k_\nu n(r)\, dr, \tag{1}$$

where r_0 is the radius at the base of the atmosphere. If the coefficient k_ν is fairly constant throughout the atmosphere, we may write Eq. (1) in the form

$$\tau_\nu^0 = k_\nu N, \tag{2}$$

where

$$N = \int_{r_0}^{\infty} n(r)\, dr. \tag{3}$$

N is the number of absorbing atoms in a column of 1 cm^2 cross section rising above the photosphere.

Let $I_0(\vartheta)$ be the intensity of radiation emerging from the photosphere into the atmosphere at an angle ϑ to the normal. Since we are only concerned with the radiation inside the absorption line, i.e., in a very small frequency interval, $I_0(\vartheta)$ can be taken constant at all the relevant frequencies. $I_0(\vartheta)$ may be assumed equal to the radiation intensity in the continuous spectrum near the particular absorption line.

Our problem is to find the intensity $I_\nu(\vartheta)$ of the radiation of frequency ν inside an absorption line leaving the atmosphere at an angle ϑ to the normal. $I_\nu(\vartheta)$ is generally expressed in units of $I_0(\vartheta)$, as the ratio

$$r_\nu(\vartheta) = \frac{I_\nu(\vartheta)}{I_0(\vartheta)} . \tag{4}$$

The ratio $r_\nu(\vartheta)$ is a shape parameter of the spectral line at an angular distance ϑ from the disk center.

The line shape in the spectrum of the entire star is also of obvious interest. To this end we define the ratio

$$r_\nu = \frac{H_\nu}{H_0} , \tag{5}$$

where H_ν is the outgoing radiation flux from the atmosphere at frequency ν inside the line, and H_0 is the incoming radiation flux from the photosphere into the atmosphere (equal to the radiation flux in the continuous spectrum near the particular line).

If the light in the stellar atmosphere were attenuated only by true absorption (i.e., due to transitions from i-th to j-th level, associated with absorption of corresponding quanta, without any reverse transitions, which emit the same quanta), the line profile could be found without any difficulty. In this case, we would obviously have

$$I_\nu(\vartheta) = I_0(\vartheta) e^{-k_\nu N \sec\vartheta} . \tag{6}$$

However, in a real atmosphere, the absorption processes are much more complex. The quanta absorbed by an atom inside a line may also be re-emitted in the same line. The emitted quanta may then be absorbed by other atoms, and then again emitted. In other words, the radiation inside the spectral line is subjected to scattering in the stellar atmosphere. Other processes (e.g., fluorescence, collisions) are also possible. As a result, the stellar atmosphere definitely emits a certain quantity of energy in the particular spectral line. Therefore, in Eq. (6), which describes the intensity of radiation transmitted by the atmosphere at a frequency ν inside a spectral line, we should introduce a further term accounting for the intensity emitted by the atmosphere at the same frequency. It is this emission contribution that constitutes the main problem of the theory of line spectra.

As the general problem of the formation of absorption lines in stellar spectra is highly complex, certain simplifying assumptions are introduced. The common simplifications are the following:

1. The radiation absorbed by a volume element in a particular spectral line can be emitted in the same line only. This rules out redistribution of energy between different spectral lines and other processes capable of destroying or creating quanta of appropriate frequency. In other words,

the stellar atmosphere is characterized by pure scattering of radiation in each spectral line.

2. The radiation absorbed by a volume element at a particular frequency in a spectral line can be emitted only at the same frequency. Redistribution of radiation over frequencies inside the line is thus also ruled out. This is the case of coherent scattering.

It follows from the above assumptions of pure and coherent scattering that the energy emitted by each volume element at a particular frequency ν is precisely equal to the energy that the element has absorbed. In other words, the stellar atmosphere is maintained in a state of monochromatic radiative equilibrium.

To arrive at a mathematical formulation of the problem, we should write the equation of transfer and the equation of radiative equilibrim for a particular line. As with the photosphere, we assume that the atmospheric layers are plane-parallel. In this case the equation of transfer takes the form

$$\cos\vartheta \frac{dI_\nu}{dr} = -\sigma_\nu I_\nu + \varepsilon_\nu, \tag{7}$$

where σ_ν is the volume absorption coefficient ($\sigma_\nu = nk_\nu$) and ε_ν is the volume emission coefficient. The condition of monochromatic radiative equilibrium is written in the form

$$4\pi\varepsilon_\nu = \sigma_\nu \int I_\nu \, d\omega, \tag{8}$$

where the integration is over the solid angles.

The optical thickness at the frequency ν is defined by

$$\tau_\nu = \int_r^\infty \sigma_\nu dr \tag{9}$$

and let

$$B_\nu = \frac{\varepsilon_\nu}{\sigma_\nu} . \tag{10}$$

Eqs. (7) and (8) thus give

$$\left.\begin{aligned} \cos\vartheta \frac{dI_\nu}{d\tau_\nu} &= I_\nu - B_\nu, \\ B_\nu &= \int I_\nu \frac{d\omega}{4\pi} . \end{aligned}\right\} \tag{11}$$

Note that Eqs. (11) are formally identical to Eqs. (10) of the preceding chapter. In the photosphere, however, these equations apply to integrated radiation, whereas in the atmosphere they correspond to a particular frequency ν inside a spectral line.

Eqs. (11) should be considered with appropriate boundary conditions. The condition at the top of the atmosphere ($\tau_\nu = 0$) indicates that the incoming radiation flux is zero:

$$I_\nu(0, \vartheta) = 0 \quad \text{for} \quad \vartheta > \frac{\pi}{2} . \tag{12}$$

The condition at the bottom of the atmosphere ($\tau_\nu = \tau_\nu^0$) indicates that the radiation intensity is equal to the incoming radiation from the photosphere:

$$I_\nu(\tau_\nu^0, \vartheta) = I_0(\vartheta) \quad \text{for} \quad \vartheta < \frac{\pi}{2}. \tag{13}$$

Our problem is thus to solve Eqs. (11) with boundary conditions (12) and (13)

These equations can be solved by one of the methods described in §1 of the preceding chapter. We will use the first approximate method.

Let I_ν' be the average outgoing radiation intensity (upward) and I_ν'' the average incoming radiation intensity (downward). Then Eqs. (11) reduce to

$$\left.\begin{gathered} \frac{1}{2}\frac{dI_\nu'}{d\tau_\nu} = I_\nu' - B_\nu, \qquad -\frac{1}{2}\frac{dI_\nu''}{d\tau_\nu} = I_\nu'' - B_\nu, \\ B_\nu = \frac{1}{2}(I_\nu' + I_\nu''). \end{gathered}\right\} \tag{14}$$

From Eqs. (14) we find

$$I_\nu' - I_\nu'' = F_\nu, \qquad I_\nu' + I_\nu'' = 2F_\nu\tau_\nu + C_\nu, \tag{15}$$

where F_ν and C_ν are arbitrary constants.

Boundary conditions (12) and (13) take the form

$$I_\nu'' = 0 \quad \text{for} \quad \tau_\nu = 0, \qquad I_\nu' = \bar{I}_0 \quad \text{for} \quad \tau_\nu = \tau_\nu^0, \tag{16}$$

where $\bar{I}_0$ is the average intensity of the incoming radiation from the photosphere. Using (16) we find

$$C_\nu = F_\nu, \qquad F_\nu = \frac{\bar{I}_0}{1 + \tau_\nu^0}. \tag{17}$$

Since the constants are known, we can obtain an expression for the function B_ν:

$$B_\nu = \frac{\bar{I}_0}{1 + \tau_\nu^0}\left(\frac{1}{2} + \tau_\nu\right). \tag{18}$$

Inserting this expression for B_ν in the equation

$$I_\nu(0, \vartheta) = \int_0^{\tau_\nu^0} B_\nu e^{-\tau_\nu \sec\vartheta} \sec\vartheta \, d\tau_\nu + I_0(\vartheta) e^{-\tau_\nu^0 \sec\vartheta},$$

which follows from the equation of transfer and boundary condition (13), we obtain the sought function $r_\nu(\vartheta)$.

The shape function r_ν for the spectrum of the entire star is particularly simple. Since in our approximation $H_\nu = \pi F_\nu$ and $H_0 = \pi \bar{I}_0$, we obtain by the second equation in (17)

$$r_\nu = \frac{1}{1 + \tau_\nu^0}, \tag{19}$$

or, by (2),

$$r_\nu = \frac{1}{1 + k_\nu N}. \tag{20}$$

Note that $\frac{1}{1+\tau_\nu^0}$ is the fraction of the photospheric radiation transmitted by the atmosphere (in general, after multiple scatterings). $\frac{\tau_\nu^0}{1+\tau_\nu^0}$ is the fraction of this radiation reflected back into the photosphere.

The solution of Eqs. (11) by the second approximate method gives

$$r_\nu = \frac{1}{1+\frac{3}{4}k_\nu N}. \tag{21}$$

It is not much different from Eq. (20).

Eqs. (11) can also be solved by the exact methods of the theory of scattering. The problem of the formation of absorption lines in stellar spectra for the Schwarzschild—Schuster model is a particular case of the general problem of the diffusion of radiation through a medium of finite optical thickness. An exact solution of this general problem was obtained by Ambartsumyan /1/ and Chandrasekhar /2/.

§88. The equivalent line width as a function of the number of absorbing atoms. The above expression for r_ν enables us to find the total energy absorbed in a spectral line. This energy is generally measured in terms of the so-called equivalent line width, which is defined as the width of a nearby region in the continuous spectrum whose energy is equal to the energy absorbed in the line (Figure 140).

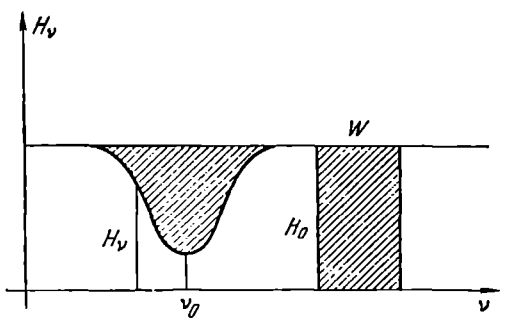


FIGURE 140

Writing W for the equivalent line width, we have by definition

$$H_0 W = \int (H_0 - H_\nu)\, d\nu. \tag{22}$$

Hence, using Eq. (5), we find

$$W = \int (1 - r_\nu)\, d\nu. \tag{23}$$

Inserting r_ν from (20), we obtain

$$W = \int \frac{k_\nu N}{1+k_\nu N}\, d\nu. \tag{24}$$

Since k_ν is presumably known, Eq. (24) relates the equivalent line width W to the number of absorbing atoms N. This relation is of the greatest significance, since it enables us to find the number of absorbing atoms from equivalent line width measurements.

The absorption coefficient k_ν is determined by a variety of factors. We will consider two only: 1) radiation damping, 2) Doppler effect due to thermal motion of atoms. In this case we have (see Vol. I, §8)

$$k_\nu = k_0 \frac{a}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-y^2}\, dy}{(p+y)^2 + a^2}. \tag{25}$$

Here p is the distance from the line center, expressed in Doppler widths, i.e.,

$$p = \frac{c}{v} \frac{\nu - \nu_0}{\nu_0}, \tag{26}$$

where ν_0 is the center frequency of the line, v is the mean thermal velocity of atoms, c is the velocity of light, a is the ratio of the natural line width to the Doppler line width, k_0 is the absorption coefficient at the center of the line for $a = 0$.

For k_0 we have

$$k_0 = \frac{\sqrt{\pi} e^2}{m \nu_0 v} f, \tag{27}$$

where m is the electron mass, e is the electron charge, f is the oscillator strength for the particular line.

Eq. (25) shows that for small p,

$$k_\nu = k_0 e^{-p^2} \qquad (p < p_0), \tag{28}$$

i.e., the absorption coefficient is determined by the Doppler effect. For large p, on the other hand,

$$k_\nu = k_0 \frac{a}{\sqrt{\pi}} \frac{1}{p^2} \qquad (p > p_0), \tag{29}$$

i.e., the main contribution is from radiation damping. The limiting value p_0 can be obtained from the equation

$$\frac{a}{\sqrt{\pi}} \frac{1}{p_0^2} = e^{-p_0^2}. \tag{30}$$

We will evaluate the integral in (24) taking k_ν from (25). Since integration in general form is impossible, we will consider three particular cases.

1. Small N. Let N be so small that $k_\nu N \ll 1$ at any frequency. Eq. (24) is thus rewritten in the form

$$W = N \int k_\nu\, d\nu. \tag{31}$$

Inserting (25) in (31), we obtain

$$W = \sqrt{\pi} \frac{\nu_0 v}{c} k_0 N. \tag{32}$$

This expression is valid only for extremely weak lines.

2. Large N. Let $k_{\nu_0}N \gg 1$, but $k_\nu N \ll 1$ in those parts of the line where k_ν is determined by radiation damping. In this case we may use expression (28) for k_ν. Inserting in (24), we get

$$W = k_0 N \frac{\nu_0 v}{c} \int_{-\infty}^{+\infty} \frac{e^{-p^2}\,dp}{1 + k_0 N e^{-p^2}}. \tag{33}$$

Approximate integration gives

$$W = 2\frac{\nu_0 v}{c}\sqrt{\lg k_0 N}. \tag{34}$$

Note that Eq. (34) can be also derived from the following considerations. We find the distance $\Delta\nu$ from the center of the line where $r_\nu = 1/2$. According to (20), at this distance we should have $k_\nu N = 1$ or

$$k_0 N e^{-\left(\frac{c}{v}\frac{\Delta\nu}{\nu_0}\right)^2} = 1. \tag{35}$$

Hence

$$\Delta\nu = \nu_0 \frac{v}{c}\sqrt{\lg k_0 N}. \tag{36}$$

Since approximately $W = 2\Delta\nu$, we return to Eq. (34).

3. Very large N. Finally let N be so large that $k_\nu N \gg 1$ at all frequencies, even far from the line center, where k_ν is determined by radiation damping. In this case, k_ν may clearly be taken in the form (29) for calculating the integral (24) over the entire line. Inserting (29) in (24) we get

$$W = \frac{a}{\sqrt{\pi}} k_0 N \frac{\nu_0 v}{c} \int_{-\infty}^{+\infty} \frac{dp}{p^2 + \frac{a}{\sqrt{\pi}} k_0 N}, \tag{37}$$

or, integrating,

$$W = \pi^{\frac{3}{4}} \frac{\nu_0 v}{c}\sqrt{a k_0 N}. \tag{38}$$

Summarizing, we see that the equivalent line width W increases with increasing number of absorbing atoms first as N, then approximately as $\sqrt{\lg N}$, and finally as $\sqrt{N}$.

In the applications a somewhat modified dependence of W on N is generally used. First, the equivalent width on the frequency scale W_ν (designated W in the preceding) is converted to the equivalent width on the wavelength scale W_λ. These two equivalent widths are related by

$$\frac{W_\lambda}{\lambda} = \frac{W_\nu}{\nu}. \tag{39}$$

Now, the number of absorbing atoms is converted to

$$X_0 = k_0 N, \tag{40}$$

which approximately corresponds to the optical thickness of the atmosphere at the center of the line (since k_{ν_0} is not greatly different from k_0 for $a \ll 1$).

The above equations can now be written in the following form:
for small X_0

$$\frac{W_\lambda}{\lambda} = \sqrt{\pi}\frac{v}{c} X_0, \tag{41}$$

for large X_0

$$\frac{W_\lambda}{\lambda} = 2\frac{v}{c}\sqrt{\lg X_0}, \tag{42}$$

for very large X_0

$$\frac{W_\lambda}{\lambda} = \frac{\pi^{\frac{1}{4}}}{2}\sqrt{\frac{v\Gamma}{cv_0} X_0}. \tag{43}$$

In the last equation we substituted

$$a = \frac{c\Gamma}{4\pi v_0 v}, \tag{44}$$

where Γ is the damping constant.

The curve which plots W as a function of N (or $\lg \frac{W_\lambda}{\lambda}$ vs. $\lg X_0$) is called the curve of growth. Growth curves are constructed using Eqs. (41), (42), and (43) and the results of numerical integration of (24) for intermediate values of X_0.

The curves of growth constitute a family which depends on two parameters: the mean thermal velocity v and the damping constant Γ. The parameters v and Γ can be replaced by the Doppler line width $\Delta\lambda_D$ and the natural line width $\Delta\lambda_E$, which are given by

$$\Delta\lambda_D = \lambda\frac{v}{c}, \quad \Delta\lambda_E = \frac{\lambda^2}{c}\frac{\Gamma}{4\pi}. \tag{45}$$

Figure 141 shows some growth curves plotted for various values of $\Delta\lambda_D$ and $\Delta\lambda_E$.

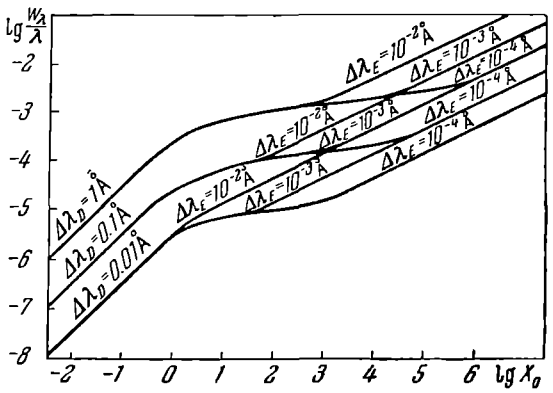


FIGURE 141

If the parameters v and Γ are known, the corresponding growth curve can be used to find X_0 from the observed equivalent line width W.

If the oscillator strength f for the given line is known, the number of absorbing atoms N can then be found from X_0.

The actual situation is more complicated, however, since in stellar atmospheres v is not only determined by thermal motion of atoms but also by the turbulent motion of the gas, whereas Γ reflects collisional damping, as well as radiation damping proper. Therefore the parameters v and Γ are never known in advance, and they should be determined separately from observations.

v and Γ are determined using curves of growth based on observational data. Growth curves can be constructed owing to the presence of multiplets in the stellar spectra. For the multiplet lines with a common lower level, N is the same and the oscillator strengths are often known. For these multiplets the values of $\lg X_0$, which according to (27) and (40) is expressed by

$$\lg X_0 = \lg f + \lg \frac{\sqrt{\pi}\, e^2}{m \nu_0 v} N, \qquad (46)$$

differ only by some unknown additive constant. By observing the equivalent line widths of the multiplet components, one can thus plot a firm growth curve, although the zero point on the horizontal axis is not certain. The corresponding sections of the growth curve can also be constructed from the lines of other multiplets. The different curve sections are then translated in the horizontal direction until they are matched, and the result is a complete curve of growth.

Figure 142 is an example of a growth curve constructed from FeI lines in the spectrum of a sunspot. This curve is borrowed from Mel'nikov and Zhuravlev /3/.

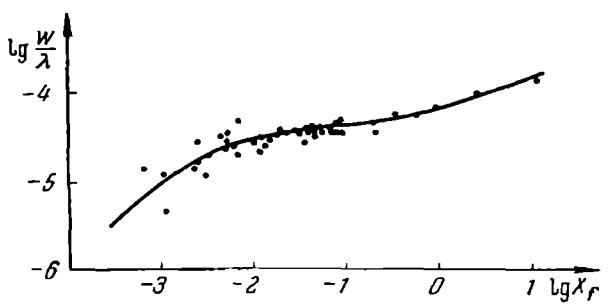


FIGURE 142

Comparison of the observed curve of growth with theoretical curves gives the parameters v and Γ for the particular star. After that the number of absorbing atoms N is readily found.

Note that joint application of growth curves and Boltzmann's equation gives the excitation temperature of the atoms in the stellar atmosphere.

Growth curves are a highly valuable tool in the study of stellar atmospheres. It suffices to note, for example, that the v of supergiants is often found to be several times greater than the mean thermal velocities of atoms.

This is associated with turbulence in stellar atmospheres. The presence of turbulence in the atmospheres of supergiants is also borne out by other data /4/.

The observed values of the parameter Γ for dwarfs are substantially higher (e.g., for the Sun, by a factor of 5—10) than the theoretical values calculated for pure radiation damping. This indicates that collisional damping is most significant in the atmospheres of dwarfs. The high values of Γ for dwarfs are associated with the relatively high density of their atmospheres.

Numerous authors have studied the atmsopheres of the Sun and stars by the method of growth curves. Considerable research has been done at the Pulkovo Observatory by O. A. Mel'nikov and co-workers.

§89. Chemical composition of stellar atmospheres. The determination of the chemical composition of stellar atmospheres is one of the fundamental problems of astrophysics. Its solution is of considerable interest, in particular, for cosmogony, since knowledge of the chemical composition of stars of various types is absolutely essential for developing a proper theory of stellar evolution.

As we have already observed, the number of atoms of a given element in a stellar atmosphere is determined from the curve of growth, which plots the equivalent line width as a function of the number of absorbing atoms. The growth curve gives the initial state population of the observed absorption line. In order to determine the total number of atoms of a given element, we require a method for converting the population of one particular state to the population of other states.

This conversion is generally effected using Boltzmannn and Saha equations, which strictly speaking are applicable only to thermodynamic equilibrium. Since the stellar atmospheres are definitely not in a state of thermodynamic equilibrium, the application of these equations leads to substantial errors. Therefore, an invariable side topic to be considered in all determinations of the chemical composition of stellar atmospheres is to what extent the actual distribution of atoms over states departs from the thermodynamic equilibrium distribution.

The Boltzmann and Saha equations are (see Vol. I, §12)

$$\frac{n_i}{n_1} = \frac{g_i}{g_1} e^{-\frac{\chi_1 - \chi_i}{kT}}, \tag{47}$$

$$n_e \frac{n^*}{n_1} = \frac{g^*}{g_1} \frac{2(2\pi m_e kT)^{\frac{3}{2}}}{h^3} e^{-\frac{\chi_1}{kT}}, \tag{48}$$

where n_i, n^* and n_e are respectively the number of neutral atoms in the i-th state, the number of ions in the ground state, and the number of free electrons in 1 cm^3, g_i is the statistical weight of the i-th state of the neutral atom, g^* is the statistical weight of the ground state of the ion, χ_i is the ionization energy from the i-th state.

Multiplying both sides of Eqs. (47) and (48) by n_1, we integrate over r from r_0 to ∞. If the temperature in the atmosphere is constant, we get

$$N_i = N_1 \frac{g_i}{g_1} e^{-\frac{\chi_1 - \chi_i}{kT}}, \tag{49}$$

$$\bar{n}_e N^+ = N_1 \frac{g^+}{g_1} \frac{2(2\pi mkT)^{\frac{3}{2}}}{h^3} e^{-\frac{\chi_1}{kT}}, \qquad (50)$$

where N_i and N^+ are respectively the number of neutral atoms in the i-th state and the number of ions in the ground state in a column of 1 cm^2 cross section above the photosphere; $\bar{n}_e$ is the average number of free electrons in 1 cm^3 in the atmosphere (it is close to the number of free electrons in 1 cm^3 near the base of the atmosphere).

If N_i (the N of the previous section) has been found from observations using the curve of growth, Eqs. (49) and (50) can be applied to determine the total number of neutral and ionized atoms of a given element (if we furthermore introduce an expression analogous to (50) for $\frac{N^{++}}{N^+}$, the total number of doubly ionized atoms can also be found, etc.).

However, the calculation of $\frac{N^+}{N_1}$ from Eq. (50) requires knowledge of the mean concentration of free electrons in the atmsophere. This quantity can be found by two methods.

The first method is applicable to stellar spectra which show lines of one element at different stages of ionization. Suppose that the spectrum contains lines of both neutral and singly ionized atoms of some element. In this case the curve of growth and Boltzmann's equation give N_1 and N^+. Eq. (50) is then applied to determine $\bar{n}_e$.

The solar spectrum contains the lines of Ca and Ca^+, and also the lines of Sr and Sr^+. Application of this method to either pair gives approximately the same result, $\bar{n}_e \approx 10^{12}$ cm^{-3}.

The second method for the determination of $\bar{n}_e$ calls for counting the number of lines of a given series which are observed in the stellar spectrum. We know that the Balmer series contains a different number of lines in the spectra of different stars. This is associated with the influence of charged particles on the high energy levels of the atom: as the result of the perturbation the outermost levels merge and become part of the energy continuum. The theory shows that the concentration of charged particles n is related to the number of the last distinguishable discrete level i by the equation /5/.

$$\lg n = 23.26 - 7.5 \lg i. \qquad (51)$$

Here $n = 2\bar{n}_e$ or $n = \bar{n}_e$, depending on the temperature.

Once $\bar{n}_e$ has been determined by either technique, we can freely use Eq. (50) to convert from N_1 to N^+ and back.

The chemical composition of stellar atmospheres has been repeatedly determined using growth curves and Eqs. (49) and (50). Table 1 lists the chemical composition of the Sun's atmsophere after Russell /6/. In this table N_1, N^+, and $N_1 + N^+$ are expressed in units of $6 \cdot 10^{12}$ atoms per 1 cm^2.

The table gives no data on the content of helium in the solar atmosphere. The reason for this is the absence of helium absorption lines in the solar spectrum. The visible spectrum may show only helium lines originating in excited states. The excitation potential of helium, however, is extremely high. Therefore, at the relatively low solar temperatures very few helium

atoms are excited and they are incapable of producing noticeable absorption lines. Strong helium absorption lines appear only in the spectra of hot stars (spectral types B and O).

TABLE 1

Atomic number	Element	$\lg N_1$	$\lg N^+$	$\lg (N_1+N^+)$	Atomic number	Element	$\lg N_1$	$\lg N^+$	$\lg (N_1+N^+)$
1	H	11.5	5.7	11.5	42	Mo	0.5	1.4	1.4
3	Li	−0.9	2.0	2.0	44	Ru	1.0	1.6	1.7
4	Be	1.8	0.8	1.8	45	Rh	−0.3	0.5	0.5
6	C	7.4	4.4	7.4	46	Pd	0.6	0.9	1.1
7	N	7.6	1.8	7.6	47	Ag	0.0	1.0	1.0
8	O	9.0	3.3	9.0	48	Cd	2.1	1.6	2.2
11	Na	4.0	7.2	7.2	49	In	−2.0	0.0	0.0
12	Mg	7.0	7.7	7.8	50	Sn	0.3	1.2	1.2
13	Al	4.6	6.4	6.4	51	Sb	0.4	0.7	0.8
14	Si	7.0	7.0	7.3	56	Ba	−0.2	3.3	3.3
16	S	5.7	3.4	5.7	57	La	−0.7	1.8	1.8
19	K	2.8	6.8	6.8	58	Ce	—	2.4	2.4
20	Ca	4.6	6.7	6.7	59	Pr	—	0.6	0.6
21	Sc	1.9	3.6	3.6	60	Nd	—	2.0	2.0
22	Ti	3.6	5.2	5.2	62	Sm	—	1.5	1.5
23	V	1.9	5.0	5.0	63	Eu	—	1.4	1.4
24	Cr	4.4	5.7	5.7	64	Gd	—	1.1	1.1
25	Mn	5.1	5.8	5.9	66	Dy	—	1.6	1.6
26	Fe	6.7	7.1	7.2	68	Er	—	0.1	0.1
27	Co	5.1	5.4	5.6	72	Hf	—	0.4	0.4
28	Ni	5.7	5.7	6.0	74	W	−0.1	−0.1	0.2
29	Cu	4.3	4.9	5.0	77	Ir	−0.5	−0.5	−0.2
30	Zn	4.9	3.8	4.9	78	Pt	1.5	1.0	1.6
31	Ga	0.2	2.0	2.0	81	Tl	−0.8	1.4	1.4
32	Ge	2.5	2.8	3.0	82	Pb	0.2	1.2	1.2
33	As	0.6	−0.7	0.6					
37	Rb	−2.5	1.7	1.7					
38	Sr	0.6	3.3	3.3					
39	Yt	0.8	2.6	2.6					
40	Zr	0.9	2.5	2.5					
41	Cb	−0.2	1.0	1.0					

The spectrum of the solar chromosphere, on the other hand, does show He emission lines. The intensity ratio of helium and hydrogen emission lines shows that the number of helium atoms is approximately 0.2 of the number of hydrogen atoms.

Hydrogen is thus the most abundant element in the solar atmosphere. Then comes helium, and after that the light elements: carbon, nitrogen, and oxygen. The number of atoms of metals, all combined, is approximately 10^{-4} of the total number of hydrogen atoms.

The chemical composition of stellar atmospheres in general is not different from the chemical composition of the Sun's atmosphere. This is approximately also the chemical composition of gaseous nebulae. The conclusion of uniform chemical compositions of stars and nebulae is of tremendous importance in cosmogony.

In some cases, however, there are significant differences in the chemical composition of stars. One of these anomalous cases are the Wolf—Rayet stars, whose spectra show the following peculiar feature: some spectra contain nitrogen emission bands but no carbon bands, whereas other spectra contain carbon bands but no nitrogen bands. This can hardly be attributed to anything else but certain differences in the chemical composition. Note that according to some authors the number of helium atoms in the

atmospheres of Wolf—Rayet stars is several times greater than the number of hydrogen atoms (in more detail this problem is discussed in Ch. XVII).

There is also a significant difference in the chemical composition in certain late-type stars. Here the spectral sequence divides into two branches: the spectra of M-type stars are characterized by TiO bands, and the spectra of N-type stars show C_2 Swan bands, CN bands, and CH bands. It is obvious that the atmospheres of M-type stars contain more oxygen than carbon, whereas the atmospheres of N-type stars are richer in carbon than in oxygen.

Other significant peculiarities in the chemical composition of N-type stars were discovered following the studies of Shain /7/ and Shain and Gaze /8/. The most remarkable of these peculiarities concerns the content of the carbon isotopes C^{12} and C^{13}. In the atmospheres of N-type stars the number of C^{13} atoms is only a few times less than the number of C^{12} atoms, whereas on Earth and in the Sun C^{13} is a rare isotope. The relative content of C^{12} and C^{13} is of particular importance, since these isotopes play an essential part in the nuclear carbon cycle.

There are possible other significant differences in the chemical composition of stars of other spectral types. For example, stars with very strong lines of some metals and stars with anomalously weak hydrogen lines have been observed. At this stage we cannot say whether this is a reflection of actual differences in the chemical composition or perhaps an outcome of peculiar excitation and ionization conditions in the stellar atmospheres.

§90. Line broadening. If the continuous radiation passes through a gas layer whose optical thickness at the line center is of the order of a few units, the width of the resulting absorption line will be of the order of the Doppler line width. As the optical thickness of the absorbing layer increases, the line width initially grows very slowly. When the optical thickness becomes very high (of the order of 10^3), the line width is seen to increase rapidly with increasing optical thickness. This is borne out by Eqs. (34) and (38) above, where the optical thickness of the layer at the line center is approximately $k_0 N$.

The Doppler line width, as we know, is a few hundredths of Å. Yet the line widths in stellar spectra often reach 1 Å and more. This means that the optical thickness of the stellar atmosphere in these lines (or, alternatively, the number of absorbing atoms) is exceedingly high.

High content of absorbing atoms in stellar atmospheres is the main reason for the broadening of absorption lines in stellar spectra. There are, however, certain other factors increasing the line width. Two of these factors are discussed in what follows.

Stark effect. Stark effect involves the splitting of atomic energy levels in an electric field. Electric fields in stellar atmospheres are set up by charged particles (electrons or ions). The charged particles move incessantly, so the electric field near a given atom is variable. The field also changes from one atom to another at any given time, owing to differences in position. The latter factor indicates that the corresponding splitting components of the energy levels of different atoms are shifted by different amounts. The line components from different atoms thus merge, so that the Stark effect produced by electrons and ions broadens the spectral line.

The theory of Stark broadening of spectral lines entails the calculation of the probabilities of various electric field values produced by different charged-particle configurations.

Let us suppose at the start that the electric field is set up only by the ion (or the electron) which is closest to the particular atom. The effect of far particles cancels out to a certain degree, and we ignore it at this stage.

If n is the number of ions in 1 cm^3, the average interionic distance x_0 is given by

$$\frac{4}{3}\pi x_0^3 n = 1, \tag{52}$$

and the probability that the nearest ion is at a distance between x and $x+dx$ from the atoms is

$$e^{-\left(\frac{x}{x_0}\right)^3} d\left(\frac{x}{x_0}\right)^3. \tag{53}$$

Let F be the field strength

$$F = \frac{e}{x^2}, \tag{54}$$

F_0 the average field strength

$$F_0 = \frac{e}{x_0^2} = \left(\frac{4\pi}{3}\right)^{\frac{2}{3}} e n^{\frac{2}{3}} = 2.60\, e n^{\frac{2}{3}}, \tag{55}$$

and $\beta = \frac{F}{F_0}$. Let $W(\beta)\,d\beta$ be the probability that β lies between β and $\beta + d\beta$. Since $\beta = \left(\frac{x_0}{x}\right)^2$, we have from (53)

$$W(\beta)\,d\beta = e^{-\beta^{-\frac{3}{2}}} d\left(-\beta^{-\frac{3}{2}}\right) = \frac{1.50}{\beta^{5/2}} e^{-\beta^{-\frac{3}{2}}} d\beta. \tag{56}$$

A more rigorous theory which allows for the combined effect of all the particles /9, 10/ gives

$$W(\beta) = \frac{2}{\pi\beta} \int_0^\infty v \sin v e^{-\left(\frac{v}{\beta}\right)^{\frac{3}{2}}} dv. \tag{57}$$

Eqs. (56) and (57) are not much different. For large β they give almost identical probabilities $W(\beta)$. This is so because high field strength is associated with a close encounter of an atom with a single ion.

Once the function $W(\beta)$ has been found, we can readily calculate the absorption coefficient in a Stark-broadened line. These calculations for a hydrogen atom are carried out as follows.

We first find the frequency distribution of the energy emitted by a volume element inside the spectral line. To this end we use the expression relating the shift $\Delta\lambda$ of the k-th splitting component to the field strength F:

$$\Delta\lambda = \frac{3h}{8\pi^2 mce} \lambda^2 n_k F, \tag{58}$$

where n_h is an integer, which depends on the initial and the final terms and the intensity distribution over the components; this number emerges from the quantum mechanical treatment of the effect. Now the absorption coefficient is assumed to be proportional to the energy emitted by the volume element and is taken to satisfy the integral relation

$$\int k_\nu \, d\nu = \frac{\pi e^2}{mc} f, \tag{59}$$

which is valid irrespective of the particular reason for line broadening (see Vol. I, §8). From this condition the absorption coefficient can be found.

The expression for the absorption coefficient far from the line center is of particular interest. Appropriate calculations give for the absorption coefficient at the wings of the Balmer lines

$$k_\lambda = C \frac{F_0^{\frac{3}{2}}}{(\lambda - \lambda_0)^{5/2}}, \tag{60}$$

where the constant C is equal to 3.13×10^{-16} for H_α, 0.885×10^{-16} for H_β, 0.442×10^{-16} for H_γ and 0.309×10^{-16} for H_δ; $\lambda - \lambda_0$ is expressed in angstroms.

Inserting for F_0 in Eq. (60) its expression from (55), we see that the absorption coefficient is proportional to the charged particle concentration n. Thus, for sufficiently high n, the absorption coefficient at the line wings is expressed by Eq. (60) rather than the earlier Eq. (29). In other words, the line width in this case is determined by the Stark effect.

Stark broadening is indeed observed in stellar atmospheres, as is evident from the following two facts.

1. The equivalent widths of the Balmer lines in stellar spectra slowly decrease with increasing line width. This fits the prediction of the Stark effect theory. If the line widths were determined by radiation damping alone, they would diminish rapidly with increasing line width owing to the rapid decline of oscillator strengths.

2. The equivalent widths of the Balmer lines in the spectra of dwarfs are much greater than in the spectra of supergiants. This is attributed to the much higher density (and hence ion concentration) of the dwarf atmospheres compared to the atmospheres of supergiants (because of large gravity on the surface of the former). The hydrogen lines in the dwarf spectra are thus greatly broadened by the Stark effect. The lines are particularly wide in the spectra of white dwarfs, whose atmospheres are the densest.

Note that the equivalent widths and profiles of hydrogen lines in stellar spectra provide an indication of the surface gravity of the stars.

The high terms of helium, lithium, and other atoms are also broadened by the Stark effect.

Stellar rotation. If the star spins on its axis, the absorption lines originating in the receding part of the disk are shifted in the red direction, and the lines from the advancing part of the disk are shifted toward the violet. On the whole, the absorption line of a rotating star is wider than the same absorption line in the spectrum of a stationary star.

Note that stellar rotation affects the widths of all the lines in the star spectrum. This is the fundamental difference between the effect of stellar rotation and the Stark effect, say, which noticeably broadens only those lines which are particularly sensitive to the electric field.

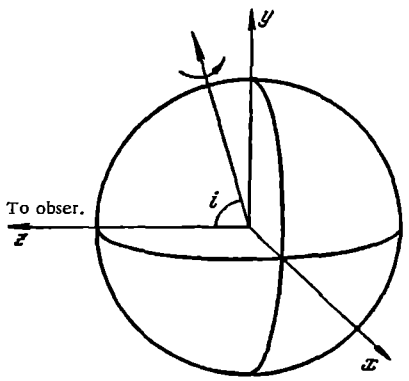


FIGURE 143

The theory of line profiles and stellar rotation was first developed by Shain and Struve /11/ and then by other authors /12/.

Let the equatorial rotation velocity of the star be v, its axis making an angle i to the line of sight. We introduce a rectangular system of coordinates x, y, z with its origin at the center of the star, the z axis directed to the observer, and the y axis lying in the plane through the rotation axis and the line of sight (Figure 143). To simplify the mathematics, we take the star's radius as unity.

Let $I(x, y, \nu - \nu_0)$ be the intensity of the radiation originating at the point x, y on the disk of a nonspinning star in a spectral line at a distance $\nu - \nu_0$ from its center. If the star is spinning, ν_0 in the expression for radiation intensity should be replaced by the central frequency for the particular point on the disk, which is

$$\nu_0 + \nu_0 \frac{v_z}{c},$$

where v_z is the radial velocity of that point. Clearly

$$v_z = -xv \sin i \tag{61}$$

Therefore the intensity of radiation of frequency ν originating at a point x, y on the disk of a rotating star is $I\left(x, y, \nu - \nu_0 + \nu_0 \frac{v}{c} x \sin i\right)$.

Let $I_0(x, y)$ be the intensity of the continuous radiation from a point x, y on the disk. Then the ratio of the energy emitted by the star at the frequency ν in the line to the continuous spectrum energy (radiated in unit frequency interval) is

$$\bar{r}(\nu - \nu_0) = \frac{\int_{-1}^{+1} dx \int_0^{\sqrt{1-x^2}} I\left(x, y, \nu - \nu_0 + \nu_0 \frac{v}{c} x \sin i\right) dy}{\int_{-1}^{+1} dx \int_0^{\sqrt{1-x^2}} I_0(x, y)\, dy}. \tag{62}$$

This expression defines the absorption line profile in the spectrum of a rotating star.

As the spin velocity increases, the absorption line becomes broader. At the same time, however, the line grows progressively shallower, since

its equivalent width is not affected: at any rotation velocity, it is equal to the equivalent width of the line in the spectrum of a nonspinning star. This result, self-evident from physical considerations, easily emerges from Eq. (62).

To simplify Eq. (62), we assume that the line profile is constant over the entire disk of the stationary star, i.e.,

$$I(x, y, \nu-\nu_0) = r(\nu-\nu_0)\, I_0(x, y). \tag{63}$$

Inserting (63) in (62), we get

$$r(\nu-\nu_0) = \int_{-1}^{+1} r\left(\nu-\nu_0+\nu_0\frac{v}{c}x\sin i\right) A(x)\,dx, \tag{64}$$

where

$$A(x) = \frac{\int_0^{\sqrt{1-x^2}} I_0(x, y)\,dy}{\int_{-1}^{+1} dx \int_0^{\sqrt{1-x^2}} I_0(x, y)\,dy}. \tag{65}$$

The distance from the line center can be expressed in units of the maximum Doppler width associated with rotation, i.e., we take

$$t = \frac{\nu-\nu_0}{\nu_0}\frac{c}{v\sin i} \tag{66}$$

and simply write $r(t)$ and $\bar{r}(t)$ for $r(\nu-\nu_0)$ and $\bar{r}(\nu-\nu_0)$. Eq. (64) is then replaced by

$$r(t) = \int_{-1}^{+1} r(t-x)\, A(x)\,dx. \tag{67}$$

Eq. (67) makes it possible to calculate the line profile in the spectrum of a rotating star from the line profile in the spectrum of a nonspinning star if $A(x)$ is known. To determine this function, we should know the intensity distribution of the continuous-spectrum radiation over the stellar disk. We take, as is usually done,

$$I_0 = C(1+\beta\cos\vartheta), \tag{68}$$

where ϑ is the angular distance from the disk center. Since $\sin\vartheta = \sqrt{x^2+y^2}$, we write

$$I_0(x, y) = C\left[1+\beta\sqrt{1-(x^2+y^2)}\right]. \tag{69}$$

Inserting (69) in (65) and integrating, we get

$$A(x) = \frac{\frac{2}{\pi}\sqrt{1-x^2}+\frac{\beta}{2}(1-x^2)}{1+\frac{2}{3}\beta}. \tag{70}$$

$A(x)$ clearly determines the line shape in the spectrum of a rotating star if the line width in the spectrum of a nonrotating star is very small. If however, this line width is not small (i.e., comparable with the width of the rotation-broadened line), the line shape should be determined using Eq. (67) and (70).

The study of stellar spectra shows that in many cases the absorption lines are broadened by rotation. In accordance with the preceding, rotation broadening is distinguished from other broadening effects by the following: 1) rotation broadens all lines without exception, 2) the width of a rotation broadened line is approximately proportional to wavelength (because of Doppler effect), 3) the equivalent line widths are independent of rotation broadening.

Using the preceding theory we can find the rotation velocity of a star (more precisely, the component $v \sin i$) from the line profile. To this end, one should take the corresponding line profile in the spectrum of a nonrotating star of the same spectral type and using Eq. (67) plot the profiles of rotation-broadened lines for various values of the parameter $v \sin i$. Comparison of these theoretical line shapes with the observed shape in the spectrum of a rotating star gives the sought value of $v \sin i$.

Figure 144 gives the shape of the 4026 Å line in the spectrum of a nonspinning star i Her and the shapes of rotation-broadened lines calculated from Eq. (67).

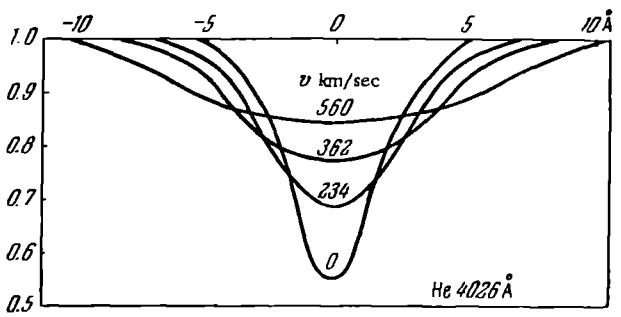


FIGURE 144

Note that firm determination of $v \sin i$ is possible only if it is of the order of a few tens of kilometers per second. Otherwise, the rotation effect is difficult to separate from other line broadening effects.

This method was actually applied to determine the value of $v \sin i$ for numerous stars. The fastest spinning stars are those of early spectral types. For example, for B stars $v \sin i$ reaches 200—300 km/sec, and in Be stars it is even higher.

The spin v of an isolated star obviously cannot be determined from its line profiles. For a large group of stars, however, $v \sin i$ can be converted to v by a statistical method. The corresponding results are of considerable significance in cosmogony.

§91. Eddington's model. The Schwarzschild-Schuster model, which divides the outer layers of a star into two components, the photosphere and

the atmosphere, is much too crude. Although in the mean the line spectrum indeed originates in the topmost layers of the star and the continuous spectrum in its deeper lying parts, there hardly seems to be a sharp boundary between them.

We will now adopt a more sophisticated model, advanced by Eddington /14/, in which each unit volume is regarded as absorbing and emitting energy in both the continuous and the line spectrum.

Strictly speaking, the continuous and the line spectrum in Eddington's model should be treated simultaneously. However, the effect of line absorption and emission on the formation of the continuous spectrum is not very pronounced. Therefore, to first approximation, when treating the continuous spectrum, we will neglect the presence of spectral lines (as in the previous model), while in the treatment of the line spectrum all the parameters characterizing the continuous spectrum are assumed to be known.

The effect of lines on the continuous spectrum was considered in a number of works /15/, but we will not go into this problem.

We will now proceed with a discussion of absorption lines in the Eddington model. Assuming a plane-parallel atmosphere, we write the equation of transfer in the form

$$\cos\vartheta \frac{dI_\nu}{dr} = -(\sigma_\nu + \alpha) I_\nu + \varepsilon_\nu + \alpha B, \tag{71}$$

where σ_ν and ε_ν are the volume absorption and emission coefficient in the spectral line, and α and αB are the volume absorption and emission coefficients in the continuous spectrum. The subscript ν is not attached to α and B, since they are not particularly sensitive to frequency inside the line.

As before, we assume that all the quanta of frequency ν in a spectral line absorbed by a volume element are emitted by the same volume element in the form of quanta of the same frequency ν. Then,

$$\varepsilon_\nu = \sigma_\nu \int I_\nu \frac{d\omega}{4\pi}. \tag{72}$$

Inserting (72) in (71) and introducing the optical depth in the continuous spectrum, τ, by means of the equality $d\tau = -\alpha\, dr$, we get

$$\cos\vartheta \frac{dI_\nu}{d\tau} = (1+\eta_\nu) I_\nu - \eta_\nu \int I_\nu \frac{d\omega}{4\pi} - B, \tag{73}$$

where

$$\eta_\nu = \frac{\sigma_\nu}{\alpha}. \tag{74}$$

In general, η_ν is a highly complex function of the optical depth, since σ_ν depends on the concentration of the atoms of a particular species in the i-th state ($\sigma_\nu = n_i k_\nu$), and α on the concentration of all atoms in various states. For simplicity we take $\eta_\nu = \text{const}$ in the atmosphere. Analysis of Eq. (73) for various particular cases of the function $\eta_\nu(\tau)$ was carried out by Eddington /14/ and Pannekoek /16/.

The $B(\tau)$ entering Eq. (73) can be assumed known from the theory of the photosphere. In particular, the following approximate expression can be used for $B(\tau)$:

$$B(\tau) = B_0(1+\beta\tau), \tag{75}$$

where B_0 and β are some constants (see Eq. (61) of the previous chapter).

We will first derive an approximate expression of Eq. (73). Let

$$\bar{I}_\nu = \int I_\nu \frac{d\omega}{4\pi}, \quad \bar{H}_\nu = \int I_\nu \cos\vartheta \frac{d\omega}{4\pi}. \tag{76}$$

$\bar{I}_\nu$ is the average radiation intensity at a point, and $4\pi\bar{H}_\nu$ is the radiation flux.

Multiplying Eq. (73) first by $\frac{d\omega}{4\pi}$ and then by $\cos\vartheta\frac{d\omega}{4\pi}$, we integrate over the solid angles to obtain

$$\frac{d\bar{H}_\nu}{d\tau} = \bar{I}_\nu - B, \tag{77}$$

$$\frac{1}{3}\frac{d\bar{I}_\nu}{d\tau} = (1+\eta_\nu)\bar{H}_\nu. \tag{78}$$

Here we used the approximate relation

$$\int I_\nu \cos^2\vartheta \frac{d\omega}{4\pi} = \frac{1}{3}\bar{I}_\nu. \tag{79}$$

From Eqs. (77) and (78) we obtain the following equation for $\bar{I}_\nu$:

$$\frac{d^2\bar{I}_\nu}{d\tau^2} = 3(1+\eta_\nu)(\bar{I}_\nu - B). \tag{80}$$

The general solution of Eq. (80) is

$$\bar{I}_\nu = C_\nu e^{-\tau\sqrt{3(1+\eta_\nu)}} + D_\nu e^{\tau\sqrt{3(1+\eta_\nu)}} + B, \tag{81}$$

where C_ν and D_ν are constants. A particular solution $I_\nu = B$ corresponds to the case when B is a linear function of τ, Eq. (75)

In the deep-lying layers of the atmsophere, where no spectral lines originate, we clearly have $I_\nu = B$. Therefore, $D_\nu = 0$. Hence,

$$\bar{I}_\nu = C_\nu e^{-\tau\sqrt{3(1+\eta_\nu)}} + B_0(1+\beta\tau). \tag{82}$$

Using (78) we get

$$\bar{H}_\nu = \frac{1}{3(1+\eta_\nu)}\left[-C_\nu e^{-\tau\sqrt{3(1+\eta_\nu)}}\sqrt{3(1+\eta_\nu)} + B_0\beta\right]. \tag{83}$$

The constant C_ν is determined from the boundary conditions. In our approximation,

$$\bar{I}_\nu = 2H_\nu \quad (\text{for } \tau = 0). \tag{84}$$

Inserting (82) and (83) in (84), we get

$$C_\nu\sqrt{3(1+\eta_\nu)} = -\frac{3(1+\eta_\nu) - 2\beta}{\sqrt{3(1+\eta_\nu)}+2} B_0. \tag{85}$$

Our aim is to determine the absorption line shape in the stellar spectrum, and we require the outgoing radiation flux, i.e., $H_\nu(0) = 4\pi\overline{H}_\nu(0)$. Taking $\tau = 0$ in Eq. (83) and using (85), we obtain

$$H_\nu(0) = 4\pi B_0 \frac{1+\frac{\beta}{\sqrt{3(1+\eta_\nu)}}}{\sqrt{3(1+\eta_\nu)}+2}. \tag{86}$$

Outside the spectral line $\eta_\nu = 0$. The radiation flux in the continuous spectrum near the spectral line is thus

$$H_0(0) = 4\pi B_0 \frac{1+\frac{\beta}{\sqrt{3}}}{\sqrt{3}+2}. \tag{87}$$

By (86) and (87)

$$r_\nu = \frac{H_\nu(0)}{H_0(0)} = \frac{1+\frac{\beta}{\sqrt{3(1+\eta_\nu)}}}{1+\frac{\beta}{\sqrt{3}}} \frac{\sqrt{3}+2}{\sqrt{3(1+\eta_\nu)}+2}. \tag{88}$$

This is the expression for the profile of an absorption line in the stellar spectrum.

Note that at the center of a strong line $\eta_\nu \gg 1$. Therefore,

$$r_\nu \approx \frac{\sqrt{3}+2}{\sqrt{3}+\beta} \frac{1}{\sqrt{\eta_\nu}}. \tag{89}$$

We see that r_ν depends on β only through the continuous radiation flux. The flux at the center of a line is virtually independent of β. This is so because the central parts of strong lines originate in the topmost layers of the atmosphere (where $B = B_0$).

At the wings of a line $\eta_\nu \ll 1$. In this case Eq. (88) yields

$$r_\nu = 1 - \frac{\eta_\nu}{2}\left(\frac{\beta}{\sqrt{3}+\beta} + \frac{\sqrt{3}}{\sqrt{3}+2}\right). \tag{90}$$

Thus $1 - r_\nu$ is proportional to η_ν or (since $\eta_\nu = \frac{nk_\nu}{\alpha}$) to the absorption coefficient inside the line, k_ν. In particular, if k_ν is determined by radiation damping, we have

$$1 - r_\lambda = \frac{c_1}{(\Delta\lambda)^2}, \tag{91}$$

and if k_ν is conditioned by the Stark effect

$$1 - r_\lambda = \frac{c_2}{(\Delta\lambda)^{5/2}}, \tag{92}$$

where c_1 and c_2 are constants.

Eq. (88) is applied to construct the "curve of growth," as in the Schwarzschild–Schuster model. The equivalent line width W is found to increase first as $k_0\frac{n}{\alpha}$, then approximately as $\sqrt{\lg k_0\frac{n}{\alpha}}$, and finally as $\sqrt{k_0\frac{n}{\alpha}}$. The growth curve in this case follows approximately the same

course as in the previous model, but now the argument of the function is the ratio $\frac{n}{\alpha}$, and not the number of absorbing atoms N. Note that physically $\frac{n}{\alpha}$ is the number of absorbing atoms in a column of 1 cm^2 cross section with a base at optical depth $\tau = 1$. Indeed, we have

$$N(\tau) = \int_r^\infty n(r)\,dr = \int_r^\infty \frac{n}{\alpha}\,\alpha\,dr = \frac{n}{\alpha}\tau, \tag{93}$$

whence the validity of our statement.

Eq. (88) defines the absorption line contour in the spectrum of the whole star. From the previous results we can find, however, the line shape $r_\nu(\vartheta)$ at an angular distance ϑ from the disk center. This requires integration of the equation of transfer (73) after substituting $\bar{I}_\nu$ from Eqs. (82)—(85). The result is

$$r_\nu(\vartheta) = \frac{1}{1+\beta\cos\vartheta}\left[1 + \frac{\beta\cos\vartheta}{1+\eta_\nu} - \frac{\eta_\nu}{1+\eta_\nu+\sqrt{3(1+\eta_\nu)}\cos\vartheta}\,\frac{3(1+\eta_\nu)-2\beta}{3(1+\eta_\nu)+2\sqrt{3(1+\eta_\nu)}}\right]. \tag{94}$$

Our solution of the problem of line shapes in star spectra, based on Eq. (79), is approximate. Nevertheless, it has been widely used in astrophysical research until quite recently. The exact solution of the problem was obtained only following the development of the general theory of radiative transfer.

Ambartsumyan /17/ applied his "invariance principle" (see §83) to derive an exact expression for $r_\nu(\zeta)$ in an isothermal atmosphere (i.e., when $B = B_0$):

$$r_\nu(\zeta) = \frac{\varphi(\zeta)}{\sqrt{1+\eta_\nu}}, \tag{95}$$

where

$$\varphi(\zeta) = 1 + \frac{\lambda}{2}\,\zeta\varphi(\zeta)\int_0^1 \frac{\varphi(\zeta')}{\zeta+\zeta'}\,d\zeta', \tag{96}$$

$$\zeta = \cos\vartheta, \quad \lambda = \frac{\eta_\nu}{1+\eta_\nu}.$$

Chandrasekhar /2/ generalized Eq. (95) to the case $B = B_0(1+\beta\tau)$. He found

$$r_\nu(\zeta) = \frac{\varphi(\zeta)}{\sqrt{1+\eta_\nu}}\left\{1 + \frac{\beta}{1+\beta\zeta}\,\frac{\eta_\nu}{1+\eta_\nu}\left[\frac{1}{2\sqrt{1+\eta_\nu}}\int_0^1 \varphi(\zeta')\,\zeta'\,d\zeta' - \zeta\right]\right\}. \tag{97}$$

Detailed tables of the function $\varphi(\zeta)$ are given in Chandrasekhar's book /2/.

§92. Central intensities of absorption lines. We know from observations that the intensity at the center of a line is a few tenths or even hundredths of the continuum intensity. Let us derive the theoretical central intensities.

For the Schwarzschild—Schuster model we have

$$r_{\nu_0} = \frac{1}{1+k_{\nu_0}N}. \tag{98}$$

To find r_{ν_0} we should first determine the number of absorbing atoms N. As we know, this number can be found from the equivalent line width W using the curve of growth. This treatment shows that for numerous lines the product $k_{\nu_0}N$, i.e., the optical thickness of the atmsophere at the center of the line, is very large, being of the order of 10^6. For these lines Eq. (98) yields $r_{\nu_0} \approx 10^{-6}$, i.e., virtually zero. In fact, however, the observed r_{ν_0} is of the order of 0.1. There is clearly an enormous divergence between theory and observation.

For Eddington's model, Eq. (89) gives

$$r_{\nu_0} \approx \left(\frac{n}{\alpha} k_{\nu_0}\right)^{-\frac{1}{2}}. \qquad (99)$$

The ratio $\frac{n}{\alpha}$, like N, can be found from the growth curve. If $\frac{n}{\alpha} k_{\nu_0} \approx 10^6$ (for constant equivalent widths W, N and $\frac{n}{\alpha}$ are approximately equal), Eq. (99) gives $r_{\nu_0} \approx 10^{-3}$. The central intensities are thus much higher than in the previous model, but they are still substantially less than the observed intensities.

The considerable divergence between the theoretical and the observed central line intensities indicates that the physical assumptions used in the two models do not correspond to the physical reality in stars. We will therefore consider the formation of absorption lines in stellar spectra without making any of the above simplifying assumptions.

We have assumed in the preceding that each quantum absorbed by a volume element is emitted by the same element in the same line. In reality, however, there is a definite redistribution of radiation between the various lines and also between the lines and the continuous spectrum.

The redistribution of radiation between spectral lines has not been studied in any detail. This process, however, cannot result in an overall increase of the central intensities of all the lines; if the intensity of some line increases, the intensities of other lines should decrease correspondingly.

The situation is different for the redistribution of radiation between the lines and the continuous spectrum. Suppose that radiation is absorbed beyond the limit of the fundamental series of a given atom. The ionized atom produced in the process eventually captures a free electron. If the electron is not captured on the first level, it emits new light quanta on jumping to lower levels. Reverse processes are naturally also possible: they involve ionization from excited states. We will show, however, that these processes make a minor contribution.

For simplicity, consider an atom with three energy states (1, 2, and 3), where the first two are discrete levels and the third is the ionization continuum. We have so far considered only scattering processes in the spectral line, i.e., transitions of the form 1→2→1. We will now introduce the following additional processes: 1) the atom is ionized from the ground state, an electron is then captured on the second level, and finally the electron drops from the second to the first level emitting a quantum inside the line (1→3→2→1 transition); 2) an atom is excited from first to second state by absorbing a line quantum, then it is ionized from the second state,

and finally captures an electron on the first level (1→2→3→1 transition). *Processes of the first kind create line quanta, and processes of the second* kind absorb line quanta.

In deep-lying atmospheric layers, where local thermodynamic equilibrium is observed, the two opposed processes mutually cancel. In the upper layers, however, the situation is different. Note that the probability of processes of the first kind depends only on the radiation density beyond the limit of the fundamental series, whereas the probability of processes of the second kind depends on the radiation density beyond the limit of the second series, as well as the radiation density in the line. The radiation density in the continuous spectrum is apparently constant throughout the atmosphere. The line radiation density, however, decreases on moving from the interior to the outside layers. Therefore in the topmost atmospheric layers processes of the first kind prevail over processes of the second kind, i.e., line quanta are created more often than they are destroyed.

Redistribution of energy between the lines and the continuous spectrum in stellar atmospheres should thus increase the central intensity of all absorption lines. To obtain a quantitative estimate of this fluorescence mechanism, we should write and solve the corresponding equation of transfer /18/.

As before, we assume Eddington's model of a stellar atmosphere and start with Eq. (71). However, Eq. (72) defining ε_ν is replaced by

$$\varepsilon_\nu = (1-|\gamma)\,\sigma_\nu \int I_\nu \frac{d\omega}{4\pi}, \tag{100}$$

where γ is the fraction of line quanta which experience true absorption (i.e., the fraction of atoms ionized from the second state). Introduction of γ accounts for processes of second kind. *

To introduce processes of the first kind, we use the volume emission coefficient ε'_ν. Eq. (71) is thus replaced by

$$\cos\vartheta \frac{dI_\nu}{dr} = -(\sigma_\nu + \alpha) I_\nu + (1-\gamma)\sigma_\nu \bar{I}_\nu + \alpha B + \varepsilon'_\nu. \tag{101}$$

An expression for ε'_ν can be obtained without difficulty. Deep in the atmosphere, where the number of first-kind processes is equal to the number of second-kind processes,

$$\varepsilon'_\nu = \gamma\sigma_\nu \bar{I}_\nu. \tag{102}$$

Also in these layers $\bar{I}_\nu = B$. Therefore Eq. (102) gives

$$\varepsilon'_\nu = \gamma\sigma_\nu B. \tag{103}$$

This expression for ε'_ν will presumably be valid in the outer layers of the atmosphere also. This is apparently so because the density of the ionizing radiation in the atmosphere hardly changes. If, however, the density of this radiation in the atmosphere may differ from the corresponding

* ε is a more conventional symbol for γ. We prefer the unconventional symbol, however, since ε is used to denote the volume emission coefficient.

radiation density in thermodynamic equilibrium, a correction factor Q is introduced in the right-hand side of Eq. (103), thus

$$\varepsilon'_\nu = Q\gamma\sigma_\nu B. \tag{104}$$

Inserting (104) in (101) and changing over from r to a new variable τ, we obtain

$$\cos\vartheta \frac{dI_\nu}{d\tau} = (1+\eta_\nu) I_\nu - (1-\gamma)\eta_\nu I_\nu - (1+Q\gamma\eta_\nu) B, \tag{105}$$

where $\eta_\nu = \frac{\sigma_\nu}{\alpha}$.

We will now derive an approximate solution of Eq. (105) taking $\eta_\nu = \text{const}$. From Eq. (105) we have

$$\frac{d\overline{H}_\nu}{d\tau} = (1+\gamma\eta_\nu)\bar{I}_\nu - (1+Q\gamma\eta_\nu) B, \tag{106}$$

$$\frac{d\bar{I}_\nu}{d\tau} = 3(1+\eta_\nu)\overline{H}_\nu. \tag{107}$$

Hence,

$$\frac{d^2\bar{I}_\nu}{d\tau^2} = 3(1+\eta_\nu)[(1+\gamma\eta_\nu)\bar{I}_\nu - (1+Q\gamma\eta_\nu) B]. \tag{108}$$

The solution of this equation is

$$\bar{I}_\nu = C_\nu e^{-\lambda_\nu\tau} + \frac{1+Q\gamma\eta_\nu}{1+\gamma\eta_\nu} B_0(1+\beta\tau), \tag{109}$$

where

$$\lambda_\nu^2 = 3(1+\eta_\nu)(1+\gamma\eta_\nu). \tag{110}$$

and C_ν is a constant. The constant before $e^{\lambda_\nu\tau}$ is zero since $\bar{I}_\nu$ must not increase exponentially with increasing τ. Inserting (109) in (107), we find

$$\overline{H}_\nu = \frac{1}{3(1+\eta_\nu)}[-\lambda_\nu C_\nu e^{-\lambda_\nu\tau} + B_0\beta]. \tag{111}$$

The constant C_ν is found from the boundary condition $\bar{I}_\nu = 2\overline{H}_\nu$ for $\tau = 0$, and this gives the following expression for the outgoing radiation flux of the star:

$$H_\nu(0) = 4\pi B_0 \frac{1+Q\gamma\eta_\nu}{1+\gamma\eta_\nu} \frac{\lambda_\nu+\beta}{3(1+\eta_\nu)+2\lambda_\nu}. \tag{112}$$

Hence,

$$r_\nu = \frac{1+Q\gamma\eta_\nu}{1+\gamma\eta_\nu} \frac{\lambda_\nu+\beta}{1+\frac{\beta}{\sqrt{3}}} \frac{\sqrt{3}+2}{3(1+\eta_\nu)+2\lambda_\nu}. \tag{113}$$

The above expression for r_ν is a generalization of Eq. (88) to the case with fluorescence.

Eq. (113) requires knowledge of γ. As we have already noted, this is the ratio of the number of ionizations from the second state to the number of ionizations plus the number of spontaneous transitions from this level. Using Einstein's transition coefficients (see Vol. I, §7), we write

$$\gamma = \frac{B_{23}\varrho_{23}}{B_{23}\varrho_{23} + A_{21}}. \tag{114}$$

Here

$$B_{23}\varrho_{23} = c\int_{\nu_{23}}^{\infty} \varrho_\nu k_{2\nu} \frac{d\nu}{h\nu}, \tag{115}$$

where ν_{23} is the frequency of radiation producing ionization from the second level, $k_{2\nu}$ is the absorption coefficient beyond the limit of the second series.

A crude estimate of γ can be obtained as follows. Suppose that $B_{23}\varrho_{23}$ is indeed a product of the radiation density immediately beyond the limit of the second series ϱ_{23} and the Einstein transition coefficient B_{23} (defined by (115)). Then writing for ϱ_{23} and A_{21}

$$\varrho_{23} = \frac{\sigma_{23}}{e^{\frac{h\nu_{23}}{kT}} - 1}, \tag{116}$$

$$A_{21} = \frac{g_1}{g_2}\sigma_{12}B_{12}, \tag{117}$$

where

$$\sigma_{ik} = \frac{8\pi h\nu_{ik}^3}{c^3} \tag{118}$$

and taking approximately $g_2 \approx g_1$, $\sigma_{12} \approx \sigma_{23}$, $B_{12} \approx B_{23}$, we obtain

$$\gamma \approx e^{-\frac{h\nu_{23}}{kT}}. \tag{119}$$

Estimating γ from Eq. (119) for atoms with an ionization potential of about 3eV from the excited state (e.g., NaI and CaI), we obtain for the temperature of the Sun $\gamma \approx 10^{-3}$. Calculations using Eqs. (114) and (115) give a figure of the same order of magnitude ($\gamma = 0.0015$ for the sodium D_1 and D_2 lines, $\gamma = 0.0004$ for the line λ 4227 Å CaI).

We will now apply Eq. (113) to the center of a line. We take

$$\gamma\eta_{\nu_0} \gg 1. \tag{120}$$

This inequality is satisfied for numerous strong lines. In these lines η_{ν_0} is very high (e.g., of the order of 10^6) and γ is often not particularly small (e.g., of the order of 10^{-3}, as above). We moreover take $Q>1$.

Under these assumptions we obtain from Eq. (113)

$$r_{\nu_0} \approx \frac{\sqrt{3}+2}{\sqrt{3}+\beta} Q\sqrt{\gamma}. \tag{121}$$

We see that Eq. (121) gives much higher values of r_{ν_0} than Eq. (89) does (it increases as inequality (120) becomes more pronounced).

Fluorescence thus substantially increases the theoretical central intensities. For $Q = 1$, however, the theoretical values of r_{ν_0} are nevertheless smaller than the observed values. For example, for the sodium D_1 and D_2 lines and the calcium line λ 4227 Å in the solar spectrum the theoretical and the observed values of r_{ν_0} differ by a factor of $2-4$. For the H and K lines of ionized calcium the discrepancy is much greater, since γ in this case is extremely low.

To make the theory consistent with observations, the hypothetical factor Q should be taken to be much less than unity. This implies that the intensity of the ultraviolet radiation of the Sun, which ionizes the atoms from the ground state, is substantially greater than the Planckian radiation intensity.

There are, however, no weighty arguments supporting this extraordinary assumption. We should therefore continue looking for additional factors affecting the central line intensities.

§93. Absorption lines with incoherent scattering. We have so far assumed that radiation quanta are scattered inside the spectral lines without change in frequency. Actually, however, various factors cause redistribution of frequencies inside the line, namely incoherent scattering. This requires further improvement of the above theory of absorption lines in stellar spectra.

The main point to be considered is to what extent, if at all, incoherent scattering is capable of increasing the central line intensities over those prescribed by the theory with coherent scattering.

We will first briefly consider the factors which cause incoherent light scattering.

1. Natural widths of the atomic energy levels. If the levels have a finite width, an atom absorbing a quantum of a certain frequency may emit a quantum of a slightly different frequency, and return after scattering to a slightly different position in the ground state. This effect is significant for lines of subordinate series, where both the upper and the lower level of each transition are fairly wide. For tne lines of the principal series, where the lower level is infinitely thin (as long as the atom is not too often excited from its ground state), the frequency of the emitted quantum coincides with the frequency of the absorbed quantum.

2. Thermal motion of atoms. Suppose that a moving atom has absorbed a quantum of a certain frequency. Since this atom may now emit a quantum in any direction, the Doppler effect may alter the frequency of the emitted quantum. The frequencies of quanta emitted by moving atoms in general are different from the frequencies of the absorbed quanta. This is so even if the light scattered by stationary atoms does not change its frequency. The redistribution of frequencies in a resonance line caused by the thermal motion of atoms was studied by Henyey /19/.

3. Effects of pressure. Let us consider the contribution of the Stark effect, which is associated with the presence of ions and free electrons in the stellar atmosphere. Suppose that at the instant when the atom absorbed a quantum it was close to a perturbing particle. During the time the atom remains in the upper state, the perturbing particle may move away, so that the energy shift changes. For this reason the frequency of the emitted quantum may be different from the frequency of the previously absorbed quantum. The energy difference between these quanta is carried away by the perturbing particle.

We will now write the equation of transfer for the case of incoherent scattering. Let $p(\nu, \nu')d\nu$ be the probability that a volume element, having absorbed a quantum of frequency ν', emits quanta of frequencies between ν and $\nu + d\nu$. The function $p(\nu, \nu')$ is determined by the combination of the above factors and in general is highly complex.

The product $p(\nu, \nu')\sigma_{\nu'}$ is the volume coefficient for absorption of quanta of frequency ν' followed by emission of quanta of frequency ν. Note that invariably

$$p(\nu, \nu')\sigma_{\nu'} = p(\nu', \nu)\sigma_{\nu}, \tag{122}$$

which corresponds to the "reciprocity principle" of the optical phenomena. Moreover,

$$\int p(\nu, \nu')d\nu = 1, \tag{123}$$

where the integration is over all the frequencies.

Using the function $p(\nu, \nu')$ we write for the volume emission coefficient

$$\varepsilon_{\nu} = \int p(\nu, \nu')\sigma_{\nu'}\,d\nu' \int I_{\nu'}\frac{d\omega}{4\pi}. \tag{124}$$

This expression for ε_{ν} should in fact be inserted in the equation of transfer (71) (in place of expression (72) for coherent scattering).

The complexity of the function $p(\nu, \nu')$ leads us to consider the limiting case of completely incoherent scattering, when the probability of scattering of a quantum of frequency ν is independent of the frequency of the absorbed quantum ν'. This is apparently the case when the line broadening is largely due to pressure effects.

If p is independent of ν', we see from (122) that $p = C\sigma_{\nu}$, where C is constant. Using (123) to determine C, we obtain

$$p = \frac{\sigma_{\nu}}{\int \sigma_{\nu'}\,d\nu'}. \tag{125}$$

Thus, in the case of completely incoherent scattering, the equation of transfer takes the form

$$\cos\vartheta\frac{dI_{\nu}}{dr} = -(\sigma_{\nu} + \alpha)I_{\nu} + \frac{\sigma_{\nu}}{\int \sigma_{\nu'}\,d\nu'}\int \sigma_{\nu'}\,d\nu' \int I_{\nu'}\frac{d\omega}{4\pi} + \alpha B. \tag{126}$$

This equation is very difficult to solve. One therefore generally uses the equality

$$\frac{\int \sigma_{\nu'}\,d\nu' \int I_{\nu'}\frac{d\omega}{4\pi}}{\int \sigma_{\nu'}\,d\nu'} = \int I_{\nu_0}\frac{d\omega}{4\pi}. \tag{127}$$

Eq. (126) then reduces to

$$\cos\vartheta\frac{dI_{\nu}}{dr} = -(\sigma_{\nu} + \alpha)I_{\nu} + \sigma_{\nu}\bar{I}_{\nu_0} + \alpha B. \tag{128}$$

Putting in Eq. (128) $\nu = \nu_0$, we can find $\bar{I}_{\nu_0}$, and then the radiation intensity $\bar{I}_\nu$. This method, however, is much too crude. In particular, it gives the same result as coherent scattering for the intensity of radiation at the line center.

The methods of the modern theory of light scattering actually yield a more exact expression for $r_\nu(\zeta)$, which corresponds to the solution of Eq. (126). Using the "invariance principle," the author /20/ obtained for an isothermal atmosphere (when $B=B_0$)

$$r_\nu(\zeta)=\frac{\eta_\nu}{1+\eta_\nu}\varphi(x)\left\{\sqrt{\int\frac{\sigma_\nu\,d\nu}{1+\eta_\nu}}-\frac{1}{2}\int_0^1\varphi(x')K_1(x')\frac{x'\,dx'}{x'+x}\right\}+\frac{1}{1+\eta_\nu}, \tag{129}$$

where the coefficient σ_ν is normalized to 1, and the function $\varphi(x)$ is obtained from the equation

$$\varphi(x)=1+\frac{x}{2}\varphi(x)\int_0^1\frac{\varphi(x')}{x+x'}K(x')\,dx' \tag{130}$$

and $x=\frac{\zeta}{1+\eta_\nu}$. The auxiliary functions $K(x)$ and $K_1(x)$ are defined by

$$K(x)=2\int_{\nu(x)}^{\infty}\eta_\nu\sigma_\nu\,d\nu,\qquad K_1(x)=2\int_{\nu(x)}^{\infty}\sigma_\nu\,d\nu, \tag{131}$$

where $\nu(x)=\nu_0$ if $x<\frac{1}{1+\eta_{\nu_0}}$, and $\frac{1}{1+\eta_{\nu(x)}}=x$ if $x>\frac{1}{1+\eta_{\nu_0}}$.

From Eq. (129) we can obtain an approximate expression for r_{ν_0} foregoing detailed calculation of the function $\varphi(x)$. It is readily seen that the second term in braces is at least half the first term. Dropping this term, and further ignoring the variation of the function $\varphi(x)$ between 0 and $\frac{1}{1+\eta_{\nu_0}}$, we find

$$r_{\nu_0}\approx\sqrt{\int\frac{\sigma_\nu\,d\nu}{1+\eta_\nu}}. \tag{132}$$

Using for the absorption coefficient inside the line the usual expression, (25), we obtain from Eq. (132)

$$r_{\nu_0}\approx\left(\frac{a}{\eta_{\nu_0}}\right)^{\frac{1}{4}}. \tag{133}$$

In case of coherent scattering, Eq. (95) approximately gives

$$r_{\nu_0}\approx\eta_{\nu_0}^{-\frac{1}{2}}. \tag{134}$$

We see that Eqs. (133) and (134) are greatly different. If, say, we take $\eta_{\nu_0}=10^6$ and $a=0.01$, we obtain for completely incoherent scattering $r_{\nu_0}\approx 0.01$ and for coherent scattering $r_{\nu_0}\approx 0.001$.

Central intensities of absorption lines for completely incoherent scattering thus can be substantially greater than in case of coherent scattering.

Eq. (129) originally derived for an isothermal atmosphere was generalized by Basbridge /21/ to the case of a linear dependence of B on the optical depth τ. Note that in this case $r_\nu(\zeta)$ is expressed in terms of the function $\varphi(x)$ defined by Eq. (130).

An exact expression for $r_\nu(\zeta)$ in the presence of fluorescence assuming a linear dependence of B on τ was derived by the author /22/. A new method was applied, based on preliminary determination of the probability of emergence of a quantum from the stellar atmosphere (in general, after multiple scattering). In this case, we approximately obtain for r_{ν_0} (under the same assumptions as in the derivation of Eq. (132))

$$r_{\nu_0} \approx \frac{1}{1+\beta\zeta} \sqrt{\left(\frac{a}{\eta_{\nu_0}}\right)^{\frac{1}{2}} + \gamma}. \tag{135}$$

In case of coherent scattering we correspondingly have

$$r_{\nu_0} \approx \frac{1}{1+\beta\zeta} \sqrt{\frac{1}{\eta_{\nu_0}} + \gamma}. \tag{136}$$

Let us compare Eqs. (135) and (136). For strong lines we definitely have

$$\left(\frac{a}{\eta_{\nu_0}}\right)^{\frac{1}{2}} \gg \frac{1}{\eta_{\nu_0}}. \tag{137}$$

Therefore if

$$\left(\frac{a}{\eta_{\nu_0}}\right)^{\frac{1}{2}} \gg \gamma, \tag{138}$$

r_{ν_0} in case of incoherent scattering is much greater than for coherent scattering. If, however,

$$\left(\frac{a}{\eta_{\nu_0}}\right)^{\frac{1}{2}} \ll \gamma, \tag{139}$$

the values of r_{ν_0} for coherent and incoherent scattering are of the same order of magnitude. In this case fluorescence predominates in the formation of the central parts of absorption lines for either scattering mechanism.

As we have already observed in the preceding, the theory of coherent scattering without fluorescences gives central intensities which are too low. Introduction of fluorescence greatly improves the results. The theoretical and the observed values of r_{ν_0} are close to one another, however, only when γ is large. If γ is small, the theory and the observations are nevertheless divergent.

The above results indicate that the incoherent scattering theory ensures a better fit with the observed central intensities of absorption lines. If γ is large, the r_{ν_0} with incoherent scattering take approximately the same values as with coherent scattering, i.e., close to the observations. If, however, γ is small, incoherent scattering ensures higher values of r_{ν_0} than coherent scattering does and thus provides better fit with observations.

Note that the difference between the different scattering modes is pronounced not only in the central intensities but also in the shapes of the absorption lines. Incoherent scattering produces absorption lines with much sharper inner parts than coherent scattering does. This difference is particularly prominent for large γ. Observations also give sharper lines than what follows from the theory of coherent scattering, although this conclusion requires further verification.

Houtgast /23/ studied the behaviour of the outer parts of absorption lines. He came to the conclusion that the variation in the wings of the strong absorption lines on the solar disk may be accounted for only by incoherent scattering.

§94. Interpretation of the spectral sequence. We have so far considered the spectrum of an individual star. Now we can discuss the entire sequence of stellar spectra.

To first approximation, the stellar spectra fall into a linear sequence. The various characteristics (e.g., the equivalent line widths) vary continuously along this sequence. This feature is associated with the dependence of stellar spectra on one principal parameter, the temperature. Variation in temperature changes the degree of excitation and ionization of atoms in the stellar atmosphere, and the line intensities are correspondingly affected.

In practice all stellar spectra are divided into distinct types in the order of decreasing temperatures: O—B—A—F—G—K—M. At the end, the spectral sequence branches into three parallel arms: besides the M-type spectra (with titanium oxide bands) there are spectra of types R—N (with carbon and cyanogen bands) and S-type spectra (with zirconium oxide bands). This branching is apparently associated with differences in the chemical composition of stars.

Let us follow the variation of the spectrum with increasing temperature. The spectra of the coldest stars (M-type, and others) show molecular bands and lines of neutral metal atoms. As the temperature increases, the molecules dissociate and the molecular bands disappear (K-type spectra). After that the metals gradually ionize. The highly complex G-type spectra contain a tremendous number of lines of neutral and ionized metals. As the temperature is further increased, the lines of the ionized metals grow in intensity (F). In spectral type A, the hydrogen Balmer lines are the strongest. In type B, helium lines appear (since the excitation of levels responsible for the helium emission lines in the visible spectrum requires sufficiently high temperatures). Finally, ionized helium lines become fairly strong in spectral type O.

The variation in the intensity of individual lines with increasing temperature can also be followed. Consider for example the lines associated with electron transitions from the excited state of a neutral atom. At low temperatures, these lines are exceedingly weak, since most atoms are in the ground state. As the temperature is raised, the degree of excitation of the atoms increases, and the equivalent widths of the corresponding lines become larger. However, the number of excited atoms increases only up to a certain temperature. As the temperature is further increased, the number of excited atoms decreases, since some are ionized. The equivalent widths of the corresponding lines therefore decrease. Thus, with

increasing stellar temperature the equivalent widths of lines associated with electron transitions from excited states of neutral atoms at first increase and then decrease.

The equivalent line widths of the ionized atoms vary in the same way (first increase and then decrease). Only the lines of the principal series of the neutral atom behave differently with increasing temperature: the equivalent widths steadily decrease (if the formation of molecules at low temperatures is ignored).

The above qualitative considerations are borne out by appropriate calculations. They are based on the use of Boltzmann and Saha equations, which determine the degree of excitation and ionization of atoms. These equations have the form

$$\frac{n_i}{n_1} = \frac{g_i}{g_1} e^{-\frac{\chi_1 - \chi_i}{kT}}, \tag{140}$$

$$n_e \frac{n^+}{n_1} = f e^{-\frac{\chi_i}{kT}}, \tag{141}$$

where

$$f = \frac{g^+}{g_1} \frac{2(2\pi m k T)^{\frac{3}{2}}}{h^3}. \tag{142}$$

Note that the application of Eqs. (140) and (141) to stellar atmospheres actually provided (in 1920) the first physical interpretation of the spectral classification /24/.

Let us apply the above equations to calculate the dependence of the equivalent line width on temperature. As in the preceding, we consider the line associated with electron transitions from the excited state of a neutral atom. If Eddington's model is used, the equivalent line width increases with increasing ratio $\frac{n_i}{\alpha}$, where n_i is the number of atoms in the i-th excited state in 1 cm^3 and α is the volume absorption coefficient in the continuous spectrum (see §91).

We write α in the form

$$\alpha = \varkappa \varrho, \tag{143}$$

where $\varkappa$ is the absorption coefficient per unit mass and ϱ is the density. Further let q be the fraction of the particular element in the total density ϱ, i.e., let

$$q\varrho = mn, \tag{144}$$

where n is the total number of atoms of the particular element in 1 cm^3 and m is the mass of one atom. Using Eqs. (143) and (144) we get

$$\frac{n_i}{\alpha} = \frac{q}{m\varkappa} \frac{n_i}{n}. \tag{145}$$

Let $n = n_1 + n^+$, i.e., we ignore the number of excited atoms and the number of doubly ionized atoms. Using Eqs. (140) and (141), we

obtain instead of (145)

$$\frac{n_i}{\alpha}=\frac{q}{m\varkappa}\frac{\frac{g_i}{g_1}e^{-\frac{\chi_i-\chi_1}{kT}}}{1+\frac{f}{n_e}e^{-\frac{\chi_1}{kT}}}. \tag{146}$$

This equation expresses the dependence of $\frac{n_i}{\alpha}$ on temperature T. Using the curve of growth, which relates the equivalent line width W to the ratio $\frac{n_i}{\alpha}$, we can find W as a function of T.

Similar equations can be derived for the lines of ionized atoms.

It follows from the preceding that the spectral type (more precisely, the equivalent line width) determines the temperature of the stellar atmosphere. The temperature determined in this way is called the ionization temperature.

Fowler and Milne /25/ proposed the following technique for the determination of ionization temperatures. Using Eq. (146), find the temperature at which $\frac{n_i}{\alpha}$ (and hence also W) is maximum. This temperature is assigned to stars of the spectral type in which the particular line actually attains its maximum equivalent width. Taking $\varkappa$ = const and $p_e=n_ekT$ = const, we obtain from Eq. (146) for the ionization temperature

$$p_e=\frac{\chi_i+\frac{5}{2}kT}{\chi_1-\chi_i}fkTe^{-\frac{\chi_1}{kT}}. \tag{147}$$

Having solved Eq. (147) (and the analogous equations for the lines of ionized atoms) for T, Fowler and Milne compared the results with the observed temperatures and derived a scale of ionization temperatures. Some of their results are listed in Table 2, where for all stars $p_e=10^{-6}$ atm.

TABLE 2

Spectral type	Maximum of the line	Ionization temperature
K5	Na, 1^2P-m^2D	3 900°
G5	Mg, 1^3P-m^3S	5 250
G0	Ca II, 1^2S-m^2P	6 290
A0	H, Balmer series	10 000
B2	He, 2^3P-m^2D	16 100
B1	Si III, OII	19 000
O5	He II, λ4686, Pickering series	35 000

However, the ioniztion temperatures obtained by this method are not very exact. The point is that the ratio $\frac{n_i}{\alpha}$ not only depends on temperature T but also on the parameters $\varkappa$ and n_e. These parameters in their turn depend on the temperature T and the gravitational acceleration g. The

equivalent line width is therefore a function of g, as well as of T. Ionization temperatures for various g values were determined by Pannekoek /26/.

The temperature required to attain a given equivalent line width was found to increase with increasing g. In the atmospheres of supergiants g is much less than in the atmospheres of dwarfs. Therefore, for a given equivalent line width, the giant temperature should be less than the dwarf temperature. In other words, the giants are cooler than the dwarfs of the same spectral type. This theoretical conclusion was confirmed by observations. The observed differences in the spectra of giants and dwarfs, however, are much more pronounced than the differences predicted by a theory using Eq. (146) and the analogous equation for $\frac{n^*}{a}$. This is largely so because the variation in g affects the equivalent line width not only through the change in the degree of ionization but also due to the changed significance of pressure effects, which directly influence the line width.

As the equivalent line width is a function of both T and g, the spectral classification requires further improvement. Each spectrum should be characterized by two parameters, related to T and g. In other words, we require a two-dimensional spectral classification. The construction of a two-dimensional spectral classification is one of the pressing problems of modern astrophysics.

The influence of the gravitational acceleration on the stellar spectrum is generally called the absolute magnitude effect. This is so because at a given temperature the gravitational acceleration g is uniquely related to the luminosity L. Indeed, we have

$$g = G\frac{M}{R^2}, \tag{148}$$

where G is the gravitational constant, M the stellar mass, R the radius, and

$$L = 4\pi R^2 \sigma T_e^4. \tag{149}$$

L and M are further related by an empirical mass—luminosity relation of the form

$$L \sim M^n, \tag{150}$$

where n is a parameter (of the order of 3—4). From the above relations we find for T_e = const

$$g \sim L^{-1+\frac{1}{n}}. \tag{151}$$

Thus, g increases as L decreases.

Another point of interest is that the effect of the gravitational acceleration on lines of neutral and ionized atoms is different. Thus, the ratio of the equivalent line widths of the ion and the neutral atoms in a stellar spectrum provides an indication of the gravitational acceleration on the star's surface and hence of its absolute magnitude. Comparison of the absolute magnitude with the stellar magnitude gives the distance of the star.

This is the basis of what is known as the method of spectroscopic parallaxes, which has yielded various valuable results during the many years of its application. In practice, certain lines of ions and neutral atoms which are particularly sensitive to the absolute magnitude effect are selected for each spectral type. The theory of this effect, however, has been insufficiently developed.

REFERENCES TO CHAPTER XV

1. AMBARTSUMYAN, V.A. Diffuse reflection of light by a turbid medium.— Doklady AN SSSR, 38, No.8. 1943.
2. CHANDRASEKHAR, S. Radiative Transfer.— 1950.
3. MEL'NIKOV, O.A. and S.S. ZHURAVLEV. Spectrophotometry of selected lines in sunspot spectra.— Vestnik, Leningrad University, No.8. 1955.
4. SHAIN, G.A. and P.F. SHAIN. Differential systematic line shifts in the spectra of white supergiants.— Izvestiya, Crimea Astrophys. Obs., 4:49. 1949.
5. MOHLER, F.L. The number of lines in a series as a function of electron pressure.— ApJ 90:429. 1939; INGLIS, D.D. and E. TELLER. Ionic depression of series limits in one-electron spectra.— ApJ 90:439. 1939.
6. RUSSELL, H.N. On the composition of the Sun's atmosphere.— ApJ 70:11. 1929.
7. SHAIN, G.A. The occurence of carbon isotopes in the spectra of N type stars.— Bull. Abastumani Obs., No.6. 1962.
8. GAZE, V.F. and G.A. SHAIN. A study of the spectrum of carbon stars.— Izvestiya, Crimea Astrophys. Obs., 2:51. 1948.
9. HOLTSMARK, J. Uber die Verbreiterung von Spektrallinien I und II.— Physik. Z., 20:162. 1919; 25:73. 1924.
10. VERWEY, S. The Stark effect of hydrogen in stellar spectra.— Diss. Amsterdam. 1936.
11. SHAIN, G. and O. STRUVE. On the rotation of the stars.— MN 89:222. 1929.
12. CARROLL, J.A. The spectroscopic determination of stellar rotation and its effects on line profiles.— MN 13:478. 1933.
13. CHANDRASEKHAR, S. and G. MUNCH. On the integral equation governing the distribution of the true and the apparent rotational velocities of stars.— ApJ 111:142. 1929.
14. EDDINGTON, A.S. The formation of absorption lines.— MN 89:620. 1929.
15. MILNE, E.A. Radiative equilibrium.—Philos. Trans. Roy. Soc. Lond. (A)223:201. 1922.
16. PANNEKOEK, A. The theoretical contours of absorption lines I and II.— MN 91:139. 519. 1930—31.
17. AMBARTSUMYAN, V.A. A turbid medium with uniform distribution of sources.— Doklady AN Armenian SSR, 8; No.4. 1948.
18. STROMGREN, B. The influence of electron capture on the contours of Fraunhofer lines.— Zeitschrift fur Astrophys., 10:237. 1935.
19. HENYEY, L.G. The Doppler effect in resonance lines.— Proc. Nat. Acad. Sci., 26:50. 1940.
20. SOBOLEV, V.V. Incoherent light scattering in stellar atmospheres.— Astron. Zh., 26:129. 1949.
21. BASBRIDGE, Ida W. Coherent and non-coherent scattering in the theory of line formation.— MN 113:52. 1953.
22. SOBOLEV, V.V. Formation of absorption lines with incoherent scattering.— Astron Zh., 31:131. 1954.
23. HOUTGAST. The Variation of the Profiles of Strong Fraunhofer Lines.— Utrecht. 1942.
24. SAHA, M.N. On a physical theory of stellar spectra.— Proc. Roy. Soc.Lond. (A), 99:134. 1921.
25. FOWLER, R.H. and E.A. MILNE.—MN 83:403. 1923.
26. PANNEKOEK, A. The stellar temperature scale.— ApJ 84:481. 1936.

V. V. SOBOLEV

Chapter XVI

GASEOUS NEBULAE

§95. The origin of nebular light. The galactic gaseous nebula fall into two groups. The first includes the so-called planetary nebulae. When viewed through a telescope, they mostly appear as circular or oval disks, reminiscent of the planetary disks, or in the form of rings. A hot star—the nucleus—is situated at the center of each planetary nebula. The second group comprises diffuse nebulae of irregular shape. Early-type stars (one or several) are observed inside the diffuse nebula or near it. The diffuse nebulae are substantially larger than the planetary nebulae.

The spectra of gaseous nebulae consists of discrete bright lines against a weak continuous background. The bright lines are identified with hydrogen, helium, ionized helium, and a number of other atoms and ions. The most characteristic feature in the spectra of gaseous nebulae are the so-called principal nebular lines N_1 and N_2 with wavelengths of 5006 and 4959 Å, respectively. Originally these lines were assigned to a new element "nebulium," not found on the Earth, but in 1923 Bowen /1/ showed that these were in fact forbidden lines of doubly ionized oxygen. The spectra of gaseous nebulae show numerous other forbidden lines too.

The currently known gaseous nebula number a few hundreds. The disks of numerous recently discovered planetary nebulae are indistinguishable through telescopes, and their assignment to planetaries is based on spectral evidence. These are either very small or very distant nebulae. The number of known diffuse nebulae greatly increased in recent years following the work Shain and Gaze /2/. Photographs of the sky taken in a narrow spectral region around the H_α line revealed faintly luminous diffuse nebulae in great numbers.

Although the energy emitted by the gaseous nebulae is concentrated predominantly in discrete spectral lines, the nebular luminosities are exceedingly high. The average absolute photographic magnitude of planetary nebulae, according to the latest data of Vorontsov-Vel'yaminov /3/, is $M_n = -0.5$. Note that planetary nebulae are mostly much brighter than their nuclei (three stellar magnitudes on the average).

The emission of the gaseous nebulae is excited by stars. The nebula abosrbs the high-frequency radiation of the star and converts it into quanta of lower frequencies, in particular visible quanta. Planetary nebulae are excited by the radiation of their nuclei (O-type and Wolf-Rayet stars). Diffuse enbulae are excited by O-type, Wolf-Rayet, and B0-type stars situated inside the nebula or near it. It is remarkable that we know of no

diffuse nebulae whose light is excited by stars of later types than B0. This is so because the intensity of the high-frequency radiation of the late-type stars is insufficient to produce marked excitation of the nebula in the visible spectrum.

The theory of emission of gaseous nebulae is one of the best developed branches of astrophysics. This is associated with the attractive simplicity of the physical processes taking place in these nebulae.

Before going into the details of this theory, let us first consider the properties of the radiation reaching a given point in the nebula from the exciting star. Suppose that the star emits as a blackbody of temperature T_*. If the sky were completely covered with these stars, the radiation density at a given point in the nebula would equal the radiation density in thermodynamic equilibrium, i.e., it would be expressed by Planck's formula

$$\varrho_\nu^* = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{\frac{h\nu}{kT_*}} - 1}. \tag{1}$$

In fact, however, the radiation density in the nebula is much less than ϱ_ν^*. It can be written in the form

$$\varrho_\nu = W\varrho_\nu^*, \tag{2}$$

where W is the so-called coefficient of dilution (attenuation) of radiation. Clearly,

$$W = \frac{\Omega}{4\pi}, \tag{3}$$

where Ω is the solid angle subtended by the star at a given point in the nebula (Figure 145). Let r_* be the radius of the star and r the distance of the particular point from its center. Since

$$\Omega = 2\pi \int_0^{\vartheta_0} \sin\vartheta \, d\vartheta = 2\pi(1 - \cos\vartheta_0),$$

and $\sin\vartheta_0 = \frac{r_*}{r}$, we get

$$W = \frac{1}{2}\left[1 - \sqrt{1 - \left(\frac{r_*}{r}\right)^2}\right]. \tag{4}$$

A point on the surface of a star collects radiation from a whole hemisphere. Therefore, in this case (i.e., for $r = r_*$), $W = 1/2$.

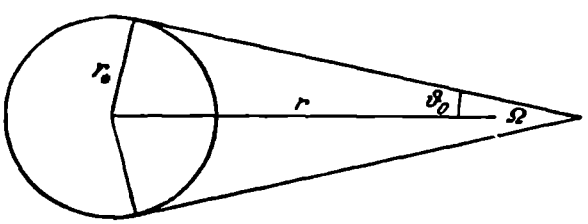


FIGURE 145

For points far from the star ($r \gg r_*$) we have by (4)

$$W = \frac{1}{4}\left(\frac{r_*}{r}\right)^2. \tag{5}$$

Note that in this case the dilution coefficient can be represented as a ratio of the stellar disk area πr_*^2 to the area of a sphere of radius r, i.e., $4\pi r^2$.

The average radius of planetary nebulae is of the order of 10^{17} cm or 10^4 astronomical units, and the radii of their nuclei are of the order of 10^{10} cm. The radiation density in a planetary nebula is therefore diluted approximately a factor of 10^{14} compared to the radiation density on the star's surface.

Integration of Eq. (2) over the frequencies, using the Stefan-Boltzmann law for the integrated radiation density in thermodynamic equilibrium, gives the following expression for the integrated radiation density in the nebula:

$$\varrho = WaT_*^4. \tag{6}$$

Writing $\varrho = aT_1^4$, we obtain

$$T_1 = W^{\frac{1}{4}} T_*. \tag{7}$$

Since the temperature of the exciting star is of the order of a few tens of thousands of degrees, while the W in nebulae is of the order of 10^{-14}, the temperature T_1 corresponding to the integrated radiation density in nebulae is of the order of a few tens of degrees.

The integrated radiation density reaching the nebula from a star is thus exceedingly small. On the other hand, we see from Eq. (2) that the relative distribution of radiation over frequencies is the same as in the original stellar radiation, i.e., it corresponds to the very high temperature T_*. The radiation reaching the nebula from the exciting star is thus characterized by an enormous disparity between the integrated density and the spectral composition.

When radiation having this property interacts with matter, the frequency distribution, as we know from thermodynamics, is changed in the direction of a more probable distribution. In other words, the high-frequency quanta are degraded to quanta of lower frequencies. This is the qualitative interpretation of the process of "degradation" of radiation in gaseous nebula.

We will now proceed with a quantitative discussion of nebular emission. At the start we assume that the atoms have only three energy states (1, 2, and 3). Of the various transitions induced by the radiation of the star, we will consider only two antagonistic cycles:

$$\text{I. } 1 \to 3 \to 2 \to 1; \quad \text{II. } 1 \to 2 \to 3 \to 1.$$

The first cycle involves absorption of one quantum of frequency ν_{13} and emission of two quanta of lower frequencies ν_{23} and ν_{12}, whereas the second cycle involves absorption of two quanta of frequencies ν_{12} and ν_{23} followed by emission of one quantum of higher frequency ν_{13}.

Let us calculate the number of events of each kind which occur in unit volume of the nebula each second. To this end we use Einstein's transition coefficients A_{ki}, B_{ik}, B_{ki} and write ϱ_{ik} for radiation density at frequency ν_{ik}.

If n_1 is the number of atoms in state 1 in 1 cm^3, the number of transitions from state 1 to state 3 in 1 cm^3 in 1 sec is $n_1 B_{13}\varrho_{13}$. From state 3 we may have transitions (both spontaneous and induced) down to state 1 and state 2. The fraction of transitions to state 2 is

$$\frac{A_{32}+B_{32}\varrho_{23}}{A_{31}+B_{31}\varrho_{13}+A_{32}+B_{32}\varrho_{23}}.$$

Some of the atoms in state 2 will return to state 3 after absorbing suitable radiation quanta, and some will drop to state 1 (spontaneously or after being induced by radiation). The ratio of the number of 2→1 transitions to the total number of transitions originating in state 2 is

$$\frac{A_{21}+B_{21}\varrho_{12}}{B_{23}\varrho_{23}+A_{21}+B_{21}\varrho_{12}}.$$

The total number of events of the first kind is thus

$$N_{\mathrm{I}} = n_1 B_{13}\varrho_{13}\,\frac{A_{32}+B_{32}\varrho_{23}}{A_{31}+B_{31}\varrho_{13}+A_{32}+B_{32}\varrho_{23}}\cdot\frac{A_{21}+B_{21}\varrho_{12}}{B_{23}\varrho_{23}+A_{21}+B_{21}\varrho_{12}}. \tag{8}$$

The number of events of the second kind is similarly found to be

$$N_{\mathrm{II}} = n_1 B_{12}\varrho_{12}\,\frac{B_{23}\varrho_{23}}{A_{21}+B_{21}\varrho_{12}+B_{23}\varrho_{23}}\cdot\frac{A_{31}+B_{31}\varrho_{13}}{A_{31}+B_{31}\varrho_{13}+A_{32}+B_{32}\varrho_{23}}. \tag{9}$$

From Eqs. (8) and (9) we obtain the following expression for the ratio of second-kind to first-kind processes:

$$\frac{N_{\mathrm{II}}}{N_{\mathrm{I}}} = \frac{B_{12}\varrho_{12}B_{23}\varrho_{23}\,(A_{31}+B_{31}\varrho_{13})}{B_{13}\varrho_{13}\,(A_{32}+B_{32}\varrho_{23})\,(A_{21}+B_{21}\varrho_{12})}. \tag{10}$$

To simplify this expression, we use Einstein's relations

$$A_{ki} = B_{ik}\frac{g_i}{g_k}\sigma_{ik}, \qquad B_{ki} = \frac{g_i}{g_k}B_{ik}, \tag{11}$$

where

$$\sigma_{ik} = \frac{8\pi h\nu_{ik}^3}{c^3}, \tag{12}$$

and g_i is the statistical weight of the i-th state. We moreover write

$$\varrho_{ik} = W\sigma_{ik}\bar{\varrho}_{ik}, \tag{13}$$

where

$$\bar{\varrho}_{ik} = \frac{1}{e^{\frac{h\nu_{ik}}{kT_*}}-1}. \tag{14}$$

Eq. (10) thus takes the form

$$\frac{N_{\mathrm{II}}}{N_{\mathrm{I}}} = W\,\frac{\bar{\varrho}_{12}\bar{\varrho}_{23}\,(1+W\bar{\varrho}_{23})}{\bar{\varrho}_{13}\,(1+W\bar{\varrho}_{12})\,(1+W\bar{\varrho}_{23})}. \tag{15}$$

For $W = 1$, Eq. (15) yields

$$\frac{N_{\mathrm{II}}}{N_{\mathrm{I}}} = \frac{\frac{1}{\bar{\varrho}_{13}}+1}{\left(\frac{1}{\bar{\varrho}_{12}}+1\right)\left(\frac{1}{\bar{\varrho}_{23}}+1\right)} = \frac{e^{\frac{h\nu_{13}}{kT_*}}}{e^{\frac{h\nu_{12}}{kT_*}}\,e^{\frac{h\nu_{23}}{kT_*}}}.$$

But $\nu_{12}+\nu_{23}=\nu_{13}$, so that in this case $\quad = 1$, as could have been expected.

If $W \ll 1$ we obtain, seeing that the factor $\frac{\bar{\varrho}_{12}\bar{\varrho}_{23}}{\bar{\varrho}_{13}}$ is of the order of unity,

$$\frac{N_{II}}{N_I} \approx W. \tag{16}$$

Thus, the number ratio of processes of the second kind to processes of the first kind is of the order of W. This result is generally known as Rosseland's theorem.

In planetary nebulae $W \approx 10^{-14}$. Therefore in this case the number of second-kind processes is negligible compared to the number of first-kind processes. In other words, the degrading of high-frequency quanta to low-frequency quanta is immeasurably more probable than the opposite process of upgrading.

§96. Determination of stellar temperatures. We have so far regarded the nebula as consisting of atoms with three energy levels only. We will now proceed with treatment of a real nebula made up of hydrogen atoms.

Because of the low radiation density in nebulae, the great majority of atoms are in the ground state. The nebulae are therefore opaque to the radiation in the Lyman series, and completely transparent to the radiation in Balmer, Paschen, and other subordinate series. The nebulae thus absorb stellar radiation at Lyman series frequencies and emit quanta in subordinate series (in particular, in the visible Balmer series), which escape unobstructed from the nebula. If the optical thickness of a nebula beyond the limit of the Lyman series is sufficiently large, the nebula will emit in hydrogen lines mainly through absorption of stellar energy in the Lyman continuum (since the energy absorbed by the nebula in Lyman lines is much less).

In more exact terms, the emission of a hydrogen nebula can be described as follows. The stellar radiation beyond the limit of the Lyman series ionizes hydrogen atoms, producing protons and free electrons. After some time, a proton captures a free electron. Suppose that an electron is captured on one of the higher lying levels. This event releases a quantum which lies beyond the limit of the corresponding subordinate series and thus escapes from the nebula. The electron then "cascades" from level to level. Owing to the exceedingly low density of radiation and matter in nebulae, these cascade transitions are hardly ever interrupted. The quanta emitted in these transitions correspond to various lines of subordinate series and also escape from the nebula. If, however, the electron drops to the first level, a Lyman line quantum is emitted, which is absorbed in the nebula and an electron is returned to the same level. Therefore, as long as the starting level is not the second, the electron will eventually end up on a level other than the first. This cascade clearly terminates with a transition to the second level with an emission of a Balmer series quantum, followed by a transition from the second level to the first with emission of a L_α quantum. The Balmer quantum escapes unobstructed from the nebula. The L_α quantum will also escape from the nebula, although after a lengthy process of diffusion.

It follows from the preceding that each absorbed Lyman continuum quantum is degraded in the nebula into precisely one Balmer quantum and one L_α quantum (some quanta in other subordinate series may also form).

In the following we assume that the optical thickness of the nebula in the Lyman continuum is substantially greater than unity. In this case, the nebula absorbs and degrades all the L_c quanta emitted by the star, so that the number of L_c quanta emitted by the star is equal to the number of Balmer quanta emitted by the nebula.

The emission of the nebula in the Balmer series thus provides an indication of the stellar emission beyond the limit of the Lyman series. Comparing the nebular glow in the Balmer series with the stellar luminosity in the visible spectrum, we essentially compare the luminosity of the star in two far removed spectral regions (ultraviolet and visible). This comparison therefore gives a tool for determining the temperature of the star.

Let I_ν^* be the average intensity of the outgoing stellar radiation. The number of quanta radiated by the star between ν and $\nu+d\nu$ is then

$$4\pi r_*^2 \frac{\pi I_\nu^*}{h\nu} d\nu,$$

so that the total number of L_c quanta emitted by the star is

$$N_{L_c}^* = 4\pi r_*^2 \int_{\nu_0}^{\infty} \frac{\pi I_\nu^*}{h\nu} d\nu, \tag{17}$$

where ν_0 is the frequency of the Lyman limit.

On the other hand, the number of Balmer quanta emitted by the nebula is

$$N_{Ba} = \sum_{Ba} \frac{E_i}{h\nu_i}, \tag{18}$$

where E_i is the total energy radiated by the nebula in the i-th Balmer line, and $h\nu_i$ is the energy of the corresponding quantum. Let E_i^* be the energy emitted by the star in a unit frequency interval around the i-th Balmer line. The nondimensional ratio

$$A_i = \frac{E_i}{\nu_i E_i^*} \tag{19}$$

can clearly be found from observation. Inserting (19) in (18) and seeing that

$$E_i^* = 4\pi r_*^2 \pi I_{\nu_i}^*,$$

we find

$$N_{Ba} = 4\pi r_*^2 \sum_{Ba} A_i \frac{\pi I_{\nu_i}^*}{h}. \tag{20}$$

If the optical thickness of the nebula beyond the Lyman limit is much greater than unity,

$$N_{Ba} = N_{L_c}^*.$$

Therefore using (17) and (20) we get

$$\sum_{\text{Ba}} A_i I^*_{\nu_i} = \int_{\nu_0}^{\infty} I^*_\nu \frac{d\nu}{\nu}. \tag{21}$$

Suppose that the radiation intensity I^*_ν is given by Planck's formula with the temperature T_*. Then, (21) is replaced by

$$\sum_{\text{Ba}} A_i \frac{\nu_i^3}{e^{\frac{h\nu_i}{kT_*}}-1} = \int_{\nu_0}^{\infty} \frac{\nu^2\, d\nu}{e^{\frac{h\nu}{kT_*}}-1}. \tag{22}$$

Substituting

$$\frac{h\nu}{kT_*} = x, \qquad \frac{h\nu_i}{kT_*} = x_i, \qquad \frac{h\nu_0}{kT_*} = x_0, \tag{23}$$

we finally get

$$\sum_{\text{Ba}} A_i \frac{x_i^3}{e^{x_i}-1} = \int_{x_0}^{\infty} \frac{x^2\, dx}{e^x-1}. \tag{24}$$

The sum in the left-hand side of this equation is taken over all the lines of the Balmer series and over the Balmer continuum.

As we have already mentioned, A_i is found from observations. After that Eq. (24) can be applied to determine the temperature T_* of the star.

This method for the determination of stellar temperatures was proposed by Zanstra /4/. He actually applied the method to find the temperatures of the nuclei of three planetary nebulae (NGC 6543, NGC 6572, and NGC 7009). The stars were found to have exceedingly high temperatures, 39,000, 40,000, and 55,000°, respectively.

Nebular emission in the lines of some other atoms (though not all) originates in the same way as the hydrogen emission through photoionization followed by recombination. These atoms absorb stellar radiation beyond the limit of their principal series and partly re-emit it in the visible spectrum. This, in particular, is the origin of the He I and He II lines of gaseous nebulae. The line intensities of these atoms can also be used to determine the stellar temperature.

Note that the stellar temperatures determined from the lines of various atoms are in general different. For example, for the nucleus of NGC 7009, hydrogen lines give 55,000° and He II lines give 70,000°. A possible explanation of this divergence is that the radiation intensity of the star is not representable by the Planckian equation with a constant temperature over the entire spectrum. It also seems that in certain cases the nebula absorbs only part of the stellar energy beyond the limit of the principal series of the particular atom. In these cases the stellar temperatures are clearly less than the true temperatures.

As we have already observed, Zanstra's mechanism of nebular emission (photoionization followed by recombination) is not the only possible one. There is an alternative mechanism operating in the gaseous nebulae which produces emission in the main nebular lines N_1 and N_2 and in other "nebulium" lines.

The different origin of the nebular emission in the N_1 and N_2 lines is proved by the following considerations.

1. Had the N_1 and N_2 quanta been produced by stellar radiation beyond the limit of the OIII principal series, the stellar temperatures would have been excessively high, sometimes over 10^6 deg.

2. Some planetary nebulae show no He II lines, as the emission of the nuclei beyond the limit of the principal series of this ion is too weak. If the N_1 and N_2 lines were a product of photo-ionization, they would also be absent in these spectra, since He^+ and O^{2+} have nearly equal ionization potentials. The N_1 and N_2 lines in the spectra of all the planetary nebulae, however, are the strongest.

The emission of gaseous nebulae in the "nebulium" lines is actually produced by excitation of atoms through collision with free electrons. The excitation potentials of states which can produce "nebulium" quanta are very high (2.5 V for N_1 and N_2). The nebulae thus contain a large number of free electrons with energies sufficient for the excitation of these states. In the final account the emission of the nebula in the "nebulium" lines is associated with stellar radiation, since the free electrons acquire their energy in photoionization.

The "nebulium" lines can also be used to determine the stellar temperatures, like the lines of recombination origin. The corresponding equations were derived by Zanstra /4/, using the following assumptions: 1) most free electrons are generated in photoionization of hydrogen atoms, 2) all the L_c quanta of the star are absorbed in the nebula, 3) all the energy acquired by the electrons in ionization goes into excitation of the "nebulium" lines.

As we know, the kinetic energy acquired by the electron which is formed when an atom is ionized by radiation of frequency ν is

$$\frac{1}{2} mv^2 = h\nu - h\nu_0,$$

where ν_0 is the ionization frequency of the atom (hydrogen in this case). If the nebula absorbs all the stellar radiation beyond the limit of the Lyman series, the energy acquired by free electrons in 1 sec is

$$4\pi r_*^2 \int_{\nu_0}^{\infty} \frac{\pi I_\nu^*}{h\nu} (h\nu - h\nu_0)\, d\nu.$$

On the other hand, the energy emitted by the nebula in "nebulium" lines in 1 sec can be expressed as

$$4\pi r_*^2 \sum_{\text{Neb}} A_i \pi I_{\nu_i}^* \nu_i,$$

where A_i are the ratios defined by Eq. (19), and the sum is taken over all the "nebulium" lines excited by electron impact.

According to Zanstra's assumptions, the last two quantities are equal:

$$\sum_{\text{Neb}} A_i I_{\nu_i}^* \nu_i = \int_{\nu_0}^{\infty} I_\nu^* (\nu - \nu_0) \frac{d\nu}{\nu}. \tag{25}$$

Substituting for I_ν^* the Planckian intensity, we get

$$\sum_{\text{Neb}} A_i \frac{\nu_i^4}{e^{\frac{h\nu_i}{kT_*}} - 1} = \int_{\nu_0}^{\infty} \frac{\nu^3 (\nu - \nu_0)}{e^{\frac{h\nu}{kT_*}} - 1} d\nu, \tag{26}$$

or using (23)

$$\sum_{\text{Neb}} A_i \frac{x_i^4}{e^{x_i}-1} = \int_{x_0}^{\infty} \frac{x^2(x-x_0)}{e^x-1}\,dx. \qquad (27)$$

Eq. (27) gives the temperature of the star, T_*, if the A_i of the "nebulium" lines are known from observations.

Using his method, Zanstra obtained 39,000° for the nucleus of NGC 6543, 38,000° for that of NGC 6552, and 50,000° for the nucleus of NGC 7009. We see that these temperatures are close to the T_* values obtained from hydrogen lines.

Various modifications of Zanstra's original method were proposed for the determination of stellar temperatures from nebular emission.

Seeing that the visual luminosity of planetary nebulae is mainly determined by the emission in the "nebulium" lines, Zanstra applied Eq. (26) to derive a relation between the star temperature T_* and the difference in the stellar magnitudes of the nucleus and the nebula m_*-m_n. Clearly, the greater the difference m_*-m_n, the higher is the temperature of the star. Using this relation, Zanstra determined the temperatures of the nuclei of 22 planetary nebulae.

Ambartsumyan /5/ derived an equation relating the stellar temperature T_* to the intensity ratio in the hydrogen H_β line and the λ 4686 Å line of ionized helium. Using this equation, he found 115,000° for the nucleus of NGC 7009 and 165,000° for the nucleus of NGC 7027. These exceedingly high temperatures (compared to the results of other methods) may be associated with the non-Planckian distribution of energy in the star's spectrum and with incomplete absorption of the L_c radiation by the nebula.

The methods described in this section for the determination of stellar temperatures are widely used in astrophysics. They can be applied not only to the nuclei of nebulae but to ordinary stars with bright spectral lines, such as Be stars, Wolf—Rayet stars, novae, etc.

§97. Ionization in nebulae. The degree of ionization of atoms in thermodynamic equilibrium is described by Saha's equation. The nebular gas, however, is not in a state of thermodynamic equilibrium, and we therefore require a new ionization equation. In this derivation we will make use of the stationarity of the nebulae. This means that the physical conditions in the nebulae do not change with time (actually, the physical conditions do change, but this variation is very slow). More precisely, we assume that the number of ionization events is equal to the number of recombination events in each volume element.

Since the atoms in nebulae are ionized mostly from the ground state, the number of ionizations in 1 cm^3 in 1 sec produced by radiation of frequencies between ν and $\nu+d\nu$ is

$$n_1 \frac{k_{1\nu}}{h\nu}\,d\nu \int I_\nu\,d\omega,$$

where $k_{1\nu}$ is the absorption coefficient per atom. Seeing that $\int I_\nu d\omega = c\rho_\nu$ and $\rho_\nu = W\rho_\nu^*$, we find for the total number of ionizations in unit volume in unit time

$$n_1 W \int_{\nu_1}^{\infty} k_{1\nu} \frac{cQ_\nu^*}{h\nu}\,d\nu,$$

where ν_1 is the frequency corresponding to ionization from the ground state.

Recombinations naturally occur to all levels. Therefore, the total number of recombinations in 1 cm^3 in 1 sec is

$$n_e n^+ \sum_1^\infty C_i(T_e),$$

where n^+ and n_e is the number of ions and electrons in 1 cm^3 and $C_i(T_e)$ is the probability that the electron is captured by the ion on the i-th level (it is a function of the electron temperature T_e).

Equating the last two expressions, we get

$$n_1 W \int_{\nu_1}^\infty k_{1\nu} \frac{c\varrho_\nu^*}{h\nu} d\nu = n_e n^+ \sum_1^\infty C_i(T_e). \tag{28}$$

This equation gives the degree of ionization of atoms in the nebula if $k_{1\nu}$ and $C_i(T_e)$ are known. It can be greatly simplified, however, if we use the relation between the coefficients $k_{i\nu}$ and $C_i(T_e)$.

Let $n_e f(v)dv$ be the number of free electrons with velocities from v to $v+dv$ in 1 cm^3 and $n_e n^+ \beta_i(v) f(v) dv$ the number of these electrons captured by ions on the i-th level in 1 cm^3 in 1 sec. Then

$$C_i(T_e) = \int_0^\infty \beta_i(v) f(v)\, dv. \tag{29}$$

On the other hand, $\beta_i(v)$ is related to the absorption coefficient $k_{i\nu}$:

$$\beta_i(v) = \frac{h^2\nu^2}{c^2 m^2 v} \frac{g_i}{g^+} k_{i\nu} \tag{30}$$

(see Vol. I, §9). For $C_i(T_e)$ we thus get

$$C_i(T_e) = \frac{h^2}{c^2 m^2} \frac{g_i}{g^+} \int_0^\infty k_{i\nu} f(v) \frac{\nu^2}{v} dv. \tag{31}$$

Eq. (28) can now be rewritten in the form

$$p n_1 W \int_{\nu_1}^\infty k_{1\nu} \frac{c\varrho_\nu^*}{h\nu} d\nu = n_e n^+ C_1(T_e), \tag{32}$$

where p is fractional capture to the first level. Using Eq. (31), and taking ρ_ν^* in the Planckian form and $f(v)$ in the Maxwellian form, we obtain from (32)

$$p n_1 W \int_{\nu_1}^\infty k_{1\nu} \frac{\nu^2 d\nu}{e^{\frac{h\nu}{kT_e}} - 1} = \frac{g_1}{g^+} n_e n^+ \frac{mh^3}{2(2\pi m k T_e)^{\frac{3}{2}}} \int_0^\infty k_{1\nu} \nu^2 e^{-\frac{mv^2}{2kT_e}} v\, dv. \tag{33}$$

Calculation of the integrals in Eq. (33) requires knowledge of $k_{1\nu}$ as a function of ν. This function is different for different atoms, but we will take approximately $k_{1\nu} \sim \frac{1}{\nu^3}$ for all atoms. The resulting error is relatively small, whereas the calculations are greatly simplified. Integrating, we obtain from Eq. (33)

$$n_e \frac{n^+}{n_1} = \frac{g^+}{g_1} pW \sqrt{\frac{T_e}{T_e}} \frac{2(2\pi m k T_e)^{\frac{3}{2}}}{h^3} \ln\left(1 - e^{-\frac{h\nu_1}{kT_e}}\right)^{-1}. \tag{34}$$

Usually we have $\frac{h\nu_1}{kT_*} \gg 1$. Therefore (34) gives

$$n_e \frac{n_+}{n_1} = \frac{g^+}{g_1} pW \sqrt{\frac{T_e}{T_*}} \frac{2(2\pi mkT_*)^{\frac{3}{2}}}{h^3} e^{-\frac{h\nu_1}{kT_*}}. \tag{35}$$

This is the final form of the ionization equation for nebulae.

We see that Eq. (35) differs from Saha's equation in that it contains a factor $pW\sqrt{\frac{T_e}{T_*}}$ in the right-hand side. This factor is very small for planetary nebulae (since $W \approx 10^{-14}$). This does not mean, however, that the degree of ionization $\frac{n^+}{n_1}$ is also small. In fact, the degree of ionization in nebulae may be quite substantial, since the small dilution coefficient W is offset by the small concentration of free electrons n_e.

Strictly speaking, Eq. (35) is valid only when the optical thickness of the nebula beyond the limit of the principal series is less than unity. Otherwise, abosrption of stellar radiation and diffuse emission by the nebula due to recombinations to the first level should be taken into consideration.

The absorption of stellar emission between the star and a given point in the nebula can be allowed for by introducing in the right-hand side of (35) a factor $e^{-\tau}$, where τ is the optical distance from the star beyond the limit of the principal series, corresponding to some averaged absorption coefficient. The ionization due to diffuse emission of the nebula can be approximately taken into consideration by dropping in the right-hand side of (28) the term which corresponds to recombinations to the first level (since in a nebula of large optical thickness the recombinations to the first level are compensated by ionizations following absorption of diffuse radiation). It is easily seen that in this case the factor p is the right-hand side of (35) is replaced by $\frac{p}{1-p}$. For the hydrogen atom the fractional capture to the first level is close to 1/2, so that the factor $\frac{p}{1-p}$ is close to unity. We will take this factor to be equal to unity for other atoms as well. We can now rewrite Eq. (35) in the form

$$n_e \frac{n^+}{n_1} = \frac{g^+}{g_1} W \sqrt{\frac{T_e}{T_*}} \frac{2(2\pi mkT_*)^{\frac{3}{2}}}{h^3} e^{-\frac{h\nu_1}{kT_*}} e^{-\tau}. \tag{36}$$

It is interesting to establish the variation in the degree of ionization $\frac{n^+}{n_1}$ with the distance r from the star. To simplify the treatment, we consider a planetary nebula whose thickness is small compared to its radius. In this case, the dilution coefficient is constant throughout the nebula (W = const). We further assume a constant concentration of atoms in the nebula (n = const).

Our calculations will strictly apply to hydrogen. In principle, however, the results are valid for all atoms which can produce substantial absorption beyond the limits of their principal series in nebulae.

Let x be the fraction of ionized atoms, i.e.,

$$n^+ = xn, \quad n_1 = (1-x)n, \quad n_e = xn. \tag{37}$$

Then, Eq. (36) is replaced by

$$\frac{x^2}{1-x} = \frac{g^+}{g_1}\frac{W}{n}\sqrt{\frac{T_e}{T_*}}\frac{2(2\pi mkT_*)^{\frac{3}{2}}}{h^3}e^{-\frac{h\nu_1}{kT_*}}e^{-\tau}. \tag{38}$$

The optical distance τ entering this equation is defined by

$$\tau = nk\int_{r_1}^{r}(1-x)\,dr, \tag{39}$$

where k is the average absorption coefficient, r_1 is the internal radius of the nebula.

Eqs. (38) and (39) readily give a differential equation between x and r. Taking the logarithm of (38) and then differentiating, we get

$$\left(\frac{2}{x}+\frac{1}{1-x}\right)dx = -d\tau. \tag{40}$$

Using (39) we find

$$\left(\frac{2}{x}+\frac{1}{1-x}\right)\frac{dx}{1-x} = -nk\,dr. \tag{41}$$

Integration of (41) yields

$$2\lg\frac{x_0}{1-x_0}\cdot\frac{1-x}{x}+\frac{1}{1-x_0}-\frac{1}{1-x} = nk(r-r_1), \tag{42}$$

where x_0 is the value of x for $\tau = 0$.

Table 1 lists as an example the values of $nk(r-r_1)$ calculated from Eq. (42) for various x. In these calculations $1-x_0 = 0.001$. The table also gives the values of τ calculated from the expression

$$\tau = \lg\left(\frac{x_0}{x}\right)^2\frac{1-x}{1-x_0}, \tag{43}$$

which follows from (38).

TABLE 1

$1-x$	$nk(r-r_1)$	τ
0.001	0	0
0.003	669	1.1
0.01	907	2.3
0.03	963	3.5
0.1	999	4.7
0.3	1009	6.4
0.5	1012	7.6

From the above equations and the table we see that x remains approximately constant up to r values determined by the relation

$$nk(r-r_1) \approx \frac{1}{1-x_0}, \tag{44}$$

and then it rapidly falls over a relatively small range of r values. The values of r from Eq. (44) correspond to τ of the order of few units.

The result is quite understandable from physical considerations. When the optical distance τ is of the order of unity, the degree of ionization decreases and the number of neutral atoms increases. The increase in the number of neutral atoms, in its turn, increases the optical distance τ.

The nebula can thus be divided into two regions: the inner region, where the degree of ionization is high $\left(\frac{n^+}{n_1} \gg 1\right)$, and the outer region, where the degree of ionization is exceedingly low $\left(\frac{n^+}{n_1} \ll 1\right)$; the boundary between these two regions is extremely sharp. The inner region emits a line spectrum of the particular element, which is produced by photoionizations and recombinations, whereas the outer region does not emit in these lines.

If the stellar temperature is sufficiently high to produce double ionization, the nebula can be divided into three shells. In the innermost region, which is the closest to the star, we find mostly doubly ionized atoms and the gas emits the line spectrum of the singly ionized atom; in the intermediate region the gas is mainly composed of singly ionized atoms and it emits the line spectrum of the neutral atom; in the outermost region the gas mostly consists of neutral elements and it does not emit in lines of recombination origin.

It follows from the preceding that the nebular radiation should display a characteristic "stratification." This theoretical conclusion is borne out by actual observations: the images of planetary nebulae obtained with a slitless spectrograph have a different size in different lines. On the whole the size of the image decreases with increasing ionization potential (as should have been expected). For example, the size of the nebula in He II lines is much less than in He I lines.

§98. Emission line intensities. The physical conditions in gaseous nebulae are so simple that the relative line intensities in the emission spectra can be found without much difficulty. If an atom in a nebula has become excited for some reason, it will mostly drop down spontaneously. In other objects (e.g., stellar atmospheres), spontaneous transitions are invariably accompanied by transitions which are induced by radiations and collisions, so that the calculation of the relative emission line intensities is a complex undertaking.

Let us first calculate the intensities of hydrogen lines in the spectra of gaseous nebulae. To this end, we use the stationarity condition for the excited hydrogen levels. This condition is an expression of the fact that the number of transitions to a particular state is equal to the number of transitions from that state.

The number of transitions to the i-th state in 1 cm^3 in 1 sec is

$$n_e n^+ C_i(T_e) + \sum_{k=i+1}^{\infty} n_k A_{ki} + n_1 B_{1i} \varrho_{1i}.$$

Here the first term is the number of captures directly to the i-th level, the second term is the number of spontaneous transitions from upperlying discrete levels, and the third term is the number of transitions from the first level induced by the Lyman line radiation.

The only transitions from the i-th state are spontaneous downward transitions. The number of these transitions in 1 cm^3 in 1 sec is

$$n_i \sum_{k=1}^{i-1} A_{ik}.$$

Equating the last two expressions, we get

$$n_i \sum_{k=1}^{i-1} A_{ik} = n_e n^+ C_i(T_e) + \sum_{k=i+1}^{\infty} n_k A_{ki} + n_1 B_{1i} \varrho_{1i} \quad (i = 2, 3, 4, \ldots). \tag{45}$$

The transitions from the first to the i-th state are induced by quanta which are emitted all over the nebula in transitions from the i-th to the first state. Since almost all the quanta are absorbed in the nebula, we have

$$n_1 B_{1i} \varrho_{1i} = n_i A_{i1}. \tag{46}$$

Using Eq. (46), we obtain from (45)

$$n_i \sum_{k=2}^{i-1} A_{ik} = n_e n^+ C_i(T_e) + \sum_{k=i+1}^{\infty} n_k A_{ki} \qquad (i = 3, 4, \ldots). \tag{47}$$

We have thus obtained a system of linear algebraic equations for the unknowns $z_k = \frac{n_k}{n_e n^+}$.

If z_k are known, the relative line intensities can easily be found. The energy emitted by the nebula in the line corresponding to the $k \to i$ transition in 1 sec is

$$E_{ki} = A_{ki} h\nu_{ik} \int n_k \, dV, \tag{48}$$

where the integral is over the entire volume of the nebula. If the electron temperature T_e, on which z_k depend, is constant throughout the nebula, Eq. (48) takes the form

$$E_{ki} = z_k A_{ki} h\nu_{ik} \int n_e n^+ \, dV. \tag{49}$$

This expression gives the line intensities, apart from a constant factor. In particular, for $i = 2$ it gives the intensity of the Balmer lines (the so-called Balmer decrement).

Eqs. (47) were solved approximately by Cillié /6/, who used the first 12 equations ($i = 3, 4, \ldots, 14$) and dropped all the rest. $C_i(T_e)$ was taken in the form

$$C_i(T_e) = \frac{2^9 \pi^5}{(6\pi)^{\frac{3}{2}}} \frac{e^{10}}{m^{\frac{3}{2}} c^3 h^3} \left(\frac{m}{kT_e}\right)^{\frac{3}{2}} \frac{1}{i^3} e^{\frac{\chi_i}{kT_e}} E_1\left(\frac{\chi_i}{kT_e}\right), \tag{50}$$

where χ_i is the energy of ionization from the i-th state.

Later Eqs. (47) were considered by Menzel and Baker /7/. They used the exact expression for $C_i(T_e)$ (with Gaunt factor different from unity) and considered higher levels as well. Their results are listed in Tables 2 and 3.

Table 2 lists the values of b_i, defined by the relation

$$n_i = b_i n_e n^+ \frac{i^2 h^3}{(2\pi m k T_e)^{\frac{3}{2}}} e^{\frac{\chi_i}{kT_e}}; \tag{51}$$

this parameter signifies to what extent the value of $\frac{n_i}{n_e n^+}$ in nebulae differs from the value of $\frac{n_i}{n_e n^+}$ in a state of thermodynamic equilibrium with temperature T_e. For $i \to \infty$, we properly have $b_i \to 1$.

TABLE 2

i \ T_e	5000°	10 000°	20 000°	40 000°	80 000°
3	0.0098	0.089	0.330	0.670	1.304
4	0.0406	0.166	0.404	0.729	1.108
5	0.0840	0.233	0.460	0.743	1.065
6	0.132	0.296	0.512	0.773	1.045
7	0.173	0.341	0.550	0.789	1.036
8	0.211	0.379	0.577	0 802	1.032
9	0.244	0.417	0.600	0.816	1.031
10	0.271	0.448	0.620	0.825	1.030
15	0.371	0.526	0.676	0.853	1.027
20	0.428	0.571	0.713	0.871	1.025
25	0.464	0.600	0.736	0.882	1.024
30	0.486	0.618	0.750	0.893	1.024

The values of the Balmer decrement calculated using these b_i values are listed in Table 3. We see that the Balmer decrement is not particularly sensitive to the electron temperature T_e.*

TABLE 3

i \ T_e	5000°	10 000°	20 000°	40 000°	80 000°	160 000°
3	2.43	2.50	2.59	2.71	2.83	2.93
4	1.00	1.00	1.00	1.00	1.00	1.00
5	0.53	0.51	0.50	0.49	0.48	0.47
6	0.33	0.31	0.30	0.29	0.27	0.26
7	0.223	0.206	0.192	0.179	0.169	0.159
8	0.157	0.143	0.130	0.120	0.112	0.104
9	0.115	0.115	0.093	0.085	0.078	0.072
10	0.087	0.079	0.069	0.062	0.057	0.052
15	0.030	0.025	0.021	0.018	0.017	0.015
20	0.014	0.012	0.010	0.009	0.006	0.006
25	0.007	0.006	0.005	0.004	0.003	0.003
30	0.004	0.003	0.003	0.002	0.002	0.001

The observed Balmer decrement is on the average steeper than the theoretical (i.e., the observed H_α/H_β ratio is greater than the theoretical figure). This divergence, however, is largely due to the reddening of distant objects because of interstellar absorption of light. If the observed Balmer decrement is corrected for interstellar absorption, a good fit is obtained with the theory.

The theoretical line intensities in the emission spectra of nebulae can also be found for other atoms. This is particularly easy for hydrogen-like ions. For an ion of atomic number Z, the energy levels are a factor of Z^2 deeper than the corresponding levels of hydrogen, the Einstein coefficients

* Recent findings indicate that for the low levels Menzel and Baker's results are less accurate than Cillié's.

of the spontaneous transitions are a factor of Z^4 larger than the corresponding hydrogen transition coefficients, and the probabilities $C_i(T_e)$ are expressed by Eq. (50) with an additional factor Z^4. Eqs. (47) and (51) therefore show that the coefficients b_i for the given ion at the temperature T_e are equal to the corresponding coefficients b_i for hydrogen at the temperature T_e/Z^2. For example, the values of b_i for ionized helium at $T_e = 40{,}000°$ can be borrowed from Table 2 at $T_e = 10{,}000°$.

The b_i of neutral helium were determined by Goldberg /8/ and Nikitin /9/. The relative HeI intensities calculated by Nikitin show good fit with observations for both the singlet and the triplet series.

§99. Forbidden lines in nebular spectra. As we have already noted, the spectra of gaseous nebulae show numerous forbidden lines. In addition to the nebular lines N_1 and N_2, identified with doubly ionized oxygen, nebular spectra also contain forbidden lines of O I, O II, N II, S II, and other elements. The strongest forbidden lines are listed in Table 4.

TABLE 4

Wavelength	Transition	Atom	Excitation potential
3726	$^4S^0_{3/2}—^2D^0_{5/2}$	O II	3.31
3729	$^4S^0_{3/2}—^2D^0_{3/2}$	O II	3.31
4363	$^1D_2—^1S_0$	O III	5.33
4959	$^3P_1—^1D_2$	O III	2.50
5007	$^3P_2—^1D_2$	O III	2.50
6583	$^3P_2—^1D_2$	N II	1.89

"Forbidden" lines differ from "allowed" lines in that their transition probabilities are extremely small. The Einstein coefficients of the spontaneous transitions for allowed lines are of the order of $10^8\,\text{sec}^{-1}$, whereas for forbidden lines they are a factor of millions and billions less. The spontaneous transition probabilities of the N_1 and N_2 lines, say, are 0.018 and 0.006 sec^{-1}, respectively.

In ordinary stellar spectra no forbidden lines are observed. We will now explain why these lines occur in nebular spectra.

First note that forbidden lines corresponding to levels from which downward transitions are permitted will never compete with the allowed lines, since for a given state the number of allowed transitions is invariably greater than the number of forbidden transitions. Forbidden lines may therefore reach comparatively high intensities only if the upper state is metastable, namely such that the only downward transitions are forbidden transitions.

However, unlike the ordinary excited states with lifetimes of the order of 10^{-8} sec, metastable states have lifetimes of a few seconds and more. Thus, for example, the average lifetime of the $^1D_2$ state of the O III ion, which emits the N_1 and N_2 lines, is 42 sec. Therefore, metastable states will emit forbidden lines only if the atom is not exposed to external distrubances for extremely long periods of time. In particular, the atom must not encounter any free electrons, since electron impact of the first kind may move the atom up from its metastable state, and electron impact of the second kind may move it down without emission in the forbidden line. Similarly, the atom should not be exposed to intense radiation, since then it may absorb a photon and jump up from its metastable state. A necessary condition for the appearance of forbidden lines in any spectrum is that the density of matter and radiation in the source be sufficiently small.

The absence of forbidden lines in stellar spectra indicates that the above condition is not met in their atmospheres. In other words,

metastable states are rapidly degraded by interaction of the excited atom with with photons or free electrons, which are present in high numbers in stellar atmospheres, and no forbidden lines are emitted. Conversely, the presence of numberous high-intensity forbidden lines in the spectra of gaseous nebulae is evidence of extremely low radiation density and density of matter in these objects.

To obtain a more exact formulation of the conditions required for the appearance of forbidden lines, let us consider an atom with three energy levels. Suppose that the forbidden transition is 2→1 (i.e., the second level is metastable), whereas downward transitions from the third level are allowed. In this case $A_{21} \ll A_{31}, A_{32}$.

Let n_2 be the number of atoms in the metastable state in unit volume. An atom may reach this state, e.g., following absorption of a quantum of frequency ν_{13} and subsequent emission of a quantum of frequency ν_{23} (i.e., in a transition 1→3→2) or as a result of an inelastic collision with free electrons.

The number of spontaneous 2→1 transitions with emission of forbidden line quanta in unit volume in unit time is $n_2 A_{21}$. At the same time there are transitions originating from the second level due to the effect of radiation and collisions which do not emit quanta in the forbidden line.

The number of induced transitions from the second level is

$$n_2 B_{21} \varrho_{12} + n_2 B_{23} \varrho_{23}.$$

Using Eqs. (11) and (13), we write this in the form

$$n_2 A_{21} W \bar{\varrho}_{12} + n_2 A_{32} \frac{g_3}{g_2} W \bar{\varrho}_{23}.$$

The first term of this expression is clearly much less than the second term (since $A_{21} \ll A_{32}$). Thus, radiation will not interfere with spontaneous transitions from the metastable state if

$$A_{21} \gg A_{32} W \bar{\varrho}_{23} \tag{52}$$

(we take here $\frac{g_3}{g_2} \approx 1$).

The number of transitions from the metastable state initiated by electron collisions is

$$n_2 n_e a_{21} + n_2 n_e b_{23}.$$

Here the first term represents transitions down from the metastable state due to second-kind electron impact, and the second term transitions up from the metastable state due to first-kind electron impact (n_e is the number of free electrons in 1 cm^3). Since first-kind electron impact can be caused only by electrons with energies higher than the excitation energy of the atom, whereas second-kind impact is caused by electrons of any energy, the first term is generally much greater than the second. Thus, collisions will not interfere with emission of quanta in the forbidden line only if

$$A_{21} \gg n_e a_{21}. \tag{53}$$

Inequalities (52) and (53) express the necessary conditions for the appearance of forbidden lines with intensities comparable to those of allowed lines.

Let us apply inequalities (52) and (53) to the "nebulium" lines N_1 and N_2. The ratio $\frac{A_{21}}{A_{32}}$ of the Einstein spontaneous-transition coefficients for the particular forbidden line and some allowed line is of the order of 10^{-10} in this case. Inequality (52) thus takes the form $W\bar{\rho}_{23} \ll 10^{-10}$. Since in planetary nebulae the dilution coefficient is of the order of 10^{-14}, this inequality is definitely satisfied.

Now, a_{21} can be written in the form $a_{21}=\bar{\sigma}_{21}\bar{v}$, where $\bar{\sigma}_{21}$ is the effective cross section for second-kind impact, $\bar{v}$ is the average electron velocity. In crude estimates we may take $\bar{\sigma}_{21} \approx 10^{-16}$ cm^2, $\bar{v} \approx 10^{-8}$ cm/sec. Therefore from (53) $n_e \ll 10^6$. As we will see in the following, $n_e \approx 10^4$ in planetary nebulae. Inequality (53) is thus also satisfied (although it is not as sharp as (52)).

Inequalities (52) and (53) can be used not only to study gaseous nebulae but also other objects, e.g., nova envelopes. When the envelope is near the star, we have the inverse inequalitities

$$A_{21} \ll A_{32}W\bar{\rho}_{23}, \quad A_{21} \ll n_e a_{21},$$

and no forbidden lines are observed. As the envelope expands, W and n_e decrease, and the necessary conditions for the appearance of forbidden lines are eventually satisfied. At the time when the forbidden line appears

$$A_{21} \approx n_e a_{21} \tag{54}$$

(at this time inequality (52) is already satisfied). The concentration n_e of free electrons in the envelope at that time can thus be found from (54).

Note that in objects with very small W and n_e there is a substantial concentration of atoms in metastable states. This clearly leads to the appearance of forbidden lines from these states. A good example is provided by the absorption line λ 3889 Å corresponding to the 2^3S metastable state of helium. This line is observed in the spectrum of the star ϑ_1 Ori in the Orion Nebula.

A detailed treatment of the necessary conditions for the appearance of forbidden lines and accumulation of atoms in metastable states was first carried out by Ambartsumyan /10/. Later this problem was also considered by other authors /11, 12/.

As we have already noted, the majority of forbidden lines in the spectra of gaseous nebulae are associated with excitation of atoms by electron impact. Therefore in all calculations connected with nebular emission in forbidden lines, we need the effective cross sections for inelastic collisions of various atoms and ions with free electrons. Recently, these effective cross sections were determined by Seaton /13/.

Let $\sigma_{ij}(v)$ be the effective cross section for the collision of an atom with a free electron in which the atom jumps from state i to state j (v is the electron velocity). This cross section may be written in the form

$$\sigma_{ij}(v) = \frac{h^2}{4\pi m^2 g_i} \frac{\Omega(i,j)}{v^2}, \tag{55}$$

where m is the electron mass, h is Planck's constant, g_i is the statistical weight of the i-th state, $\Omega(i, j)$ is the nondimensional effective cross section.

Seaton calculated the function $\Omega(i, j)$ for ions with $2p^q$ and $3p^q$ configurations ($q = 2, 3, 4$). These configurations contain three terms, in the order of increasing energy: ^3P, ^1D, and ^1S ($q = 2$ or 4) or ^4S, ^2D, and ^2P ($q = 3$), designated in what follows 1, 2, 3, respectively. The values of $\Omega(i, j)$ for ions with $2p^q$ configurations are listed in Table 5.

TABLE 5

q=2	Ω(1,2)	Ω(1,3)	q=3	Ω(1,2)	Ω(1,3)	q=4	Ω(1,2)	Ω(1,3)
NII	2.39	0.223	OII	1.44	0.218	FII	(0.95)	(0.057)
OIII	1.73	0.195	FIII	(1.00)	(0.221)	NeIII	0.76	0.077
FIV	(1.21)	(0.172)	NeIV	(0.68)	(0.234)	NaIV	(0.61)	(0.092)
NeV	(0.84)	(0.157)	NaV	(0.43)	(0.225)	MgV	(0.54)	(0.112)

For the SII ion with $3p^q$ configuration we have $\Omega(1,2)=2.02$, $\Omega(1,3)=0.383$. The calculated values of $\Omega(i, j)$ differ from the exact values apparently by no more than 40%, and the estimates (in parentheses) no more than by a factor of two.

The values of Ω for OIII listed in the table determine the cross sections for impact excitation of the $^1D_2$ state from which the N_1 and N_2 lines originate and of the $^1S_0$ state from which the λ 4363 Å line is emitted.

Knowledge of $\Omega(i, j)$ makes it possible to determine the number of impacts of first or second kind in unit volume in unit time. If the electron velocity is v, the number of inelastic electron-atom collisions inducing $i \rightarrow j$ transitions in 1 cm^3 in 1 sec is $n_i n_e \sigma_{ij}(v) v$. We assume that the electron velocities have a Maxwellian distribution

$$dn_e = 4\pi n_e \left(\frac{m}{2\pi k T_e}\right)^{\frac{3}{2}} e^{-\frac{mv^2}{2kT_e}} v^2 dv. \qquad (56)$$

where T_e is the temperature of the electron gas. In this case the total number of electron-atom collisions inducing $i \rightarrow j$ transitions is

$$n_i \int_{v_0}^{\infty} \sigma_{ij}(v) \frac{dn_e}{dv} v\, dv,$$

where v_0 is obtained from the equality $\frac{mv_0^2}{2} = h\nu_{ij}$ for impacts of the first kind and $v_0 = 0$ for second-kind impacts. As before, let $n_i n_e b_{ij}$ be the number of first-kind impacts in 1 cm^3 in 1 sec and $n_j n_e a_{ji}$ the number of second-kind impacts. Using (55) and (56) we then obtain

$$b_{ij} = \frac{\Omega(i,j)}{g_i} \frac{h^2}{2\pi m^2} \left(\frac{m}{2\pi k T_e}\right)^{\frac{1}{2}} e^{-\frac{h\nu_{ij}}{kT_e}} = 8.54 \times 10^{-6} \frac{\Omega(ij)}{g_i \sqrt{T_e}} e^{-\frac{h\nu_{ij}}{kT_e}} \qquad (57)$$

and

$$a_{ji} = 8.54 \times 10^{-6} \frac{\Omega(ij)}{g_j \sqrt{T_e}}. \qquad (58)$$

From (57) and (58) we obtain the following relation between the coefficients b_{ij} and a_{ji}:

$$b_{ij} = a_{ji} \frac{g_j}{g_i} c^{-\frac{h\nu_{ij}}{kT_e}}. \tag{59}$$

This relation is applicable whenever a Maxwellian distribution of electron velocities is valid (see Vol. I, §10).

§100. Nebular temperatures. Comparison of the above theory of nebular emission with observation results yields various parameters which characterize the physical conditions in nebulae, namely temperatures, densities, etc. Two methods have been proposed for the determination of the electron temperatures of nebulae.

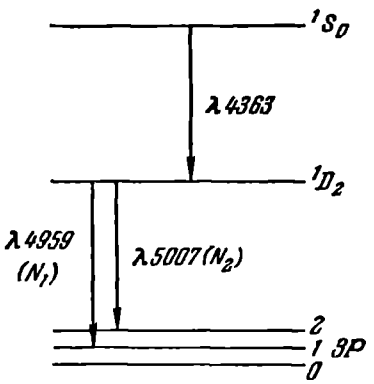


FIGURE 146

The first method is based on the intensity ratio of the forbidden lines λ 4363 Å and $N_1 + N_2$ of doubly ionized oxygen. The N_1 and N_2 lines originate in $^1D_2 \rightarrow {}^3P_1$ and $^1D_2 \rightarrow {}^3P_2$ transitions, and the λ 4363 Å line in the $^1S_0 \rightarrow {}^1D_2$ transition (Figure 146). The excitation potentials of the levels $^1D_2$ and $^1S_0$ are 2.5 and 5.3 V, respectively. These levels are excited by electron impact. The higher the electron temperature of the nebula, the greater is the ratio of the atoms in the $^1S_0$ state to the number of atoms in the $^1D_2$ state and consequently the brighter is the λ 4363 Å line compared to the N_1 and N_2.

Let n_1, n_2, n_3 stand for the number of O^{2+} ions in the ground state (^3P) and in the two metastable states ($^1D_2$ and $^1S_0$), respectively. Ignoring collisional transitions from excited states compared to spontaneous transitions, we obtain from the stationarity condition

$$\left.\begin{aligned} n_1 n_e b_{12} + n_3 A_{32} &= n_2 A_{21}, \\ n_1 n_e b_{13} &= n_3 (A_{31} + A_{32}). \end{aligned}\right\} \tag{60}$$

For the O^{2+} ion the probability of the 3→1 forbidden transition is much less than the probability of the 3→2 transition, so that $A_{31} \ll A_{32}$. By (60) we thus have

$$\frac{n_2 A_{21}}{n_3 A_{32}} = 1 + \frac{b_{12}}{b_{13}}.$$

The required line intensity ratio is therefore

$$\frac{E_{N_1+N_2}}{E_{\lambda 4363}} = \frac{\nu_{12}}{\nu_{23}} \left(1 + \frac{b_{12}}{b_{13}}\right). \tag{61}$$

Let us now change over from b_{ij} to a_{ji} using Eq. (59). This will enable us to express in explicit form the dependence of the intensity

ratio $\frac{E_{N_1+N_2}}{E_{\lambda 4363}}$ on the electron temperature T_e, since a_{ji} is hardly dependent on T_e. After appropriate substitution, we get

$$\frac{E_{N_1+N_2}}{E_{\lambda 4363}} = \frac{\nu_{12}}{\nu_{23}}\left(1+\frac{g_2}{g_3}\frac{a_{21}}{a_{31}}e^{\frac{h\cdot\nu_{31}}{kT_e}}\right). \tag{62}$$

Eq. (62) was derived by Ambartsumyan /14/. Taking $\frac{g_2a_{21}}{g_3a_{31}}$ to be of the order of unity and using the observed intensity ratio of the $N_1 + N_2$ and λ 4363 Å lines (= 30), he obtained some 7,000° for the electron temperature of the nebula.

This method of electron temperature determination of nebulae was further developed by Menzel, Aller, and Hebb /15/, who calculated the effective cross sections for the collisional excitation of nebulium lines and introduced second-kind impact in the stationarity equations. Taking the average figure $n_e \approx 10^4$ for the electron concentration, they obtained the equation

$$\frac{E_{N_1+N_2}}{E_{\lambda 4363}} = 4.07e^{\frac{33000}{T_e}} \tag{63}$$

The electron temperatures from Eq. (63) were found to be of the order of 6,000—10,000° for numerous nebulae.

Seaton's effective collision cross sections /13/ give somewhat higher values for the electron temperatures of the nebulae /16, 17/.

The second method for the determination of T_e of nebulae is based on examination of the energy balance of electron gas. In photoionization the electrons acquire a certain energy, which is imparted to them by the star's radiation. Subsequently they dissipate this energy in the excitation of atoms with which they collide and as radiation emitted in the continuous spectrum (free-free transitions and recombinations). These processes naturally establish a certain temperature of the electron gas.

Suppose that the electrons gain energy through photoionization of hydrogen atoms and lose it in three ways: 1) through radiation in the continuous spectrum, 2) through excitation of O^{2+} ions, and 3) through excitation of hydrogen atoms. Under these assumptions the author /18, 19/ derived the following relation between the star's temperature T_* and the electron temperature of the nebula T_e:

$$AT_* = BT_e + C\frac{N_2}{H_\beta} + D\frac{\bar{n}_1}{\bar{n}^+}, \tag{64}$$

where A, B, C, D are some coefficients (tabulated in /18, 19/). The first of these coefficients is a function of T_* only, and the others are a function of T_e only; N_2/H_β is the intensity ratio of the lines N_2 and H_β, $\frac{\bar{n}_1}{\bar{n}^+}$ is the average (for the entire nebula) ratio of the number of neutral hydrogen atoms to the number of protons. Each of the terms in the right-hand side of Eq. (64) corresponds to a certain energy loss mechanism of the electron gas.

Calculations using Eq. (64) gave for the temperatures of planetary nebulae values of the order of 9,000—14,000°. These low T_e values are attributable to the cooling of the electron gas through inelastic collisions with atoms (expecially with O^{2+} ions).

Recently, Aller /17/ derived an equation analogous to (64) (though without the last term) and applied it to calculate the electron temperatures of nebulae.

§101. Masses, densities, and chemical composition. The emission of nebulae in hydrogen lines can be used to determine the concentration of hydrogen atoms in nebulae. To this end one generally uses Eq. (49) which gives the energy emitted by a nebula in a given line. Applying this expression to the Balmer line corresponding to the $k \rightarrow 2$ transition, we get

$$E_{k2} = z_k A_{k2} h\nu_{2k} \int n_e n^+ \, dV. \tag{65}$$

Since hydrogen is the most abundant element, and in the emitting part of the nebula it mostly occurs in ionized state, we may take $n_e = n^+$. Eq. (65) is now written in the form

$$E_{k2} = z_k A_{k2} h\nu_{2k} n^{+2} V, \tag{66}$$

where n^+ is the average number of protons in 1 cm^3 and V is the volume of the emitting region of the nebula. By (66) we have

$$n^+ = \sqrt{\frac{E_{k2}}{z_k A_{k2} h\nu_{2k} V}}. \tag{67}$$

An estimate of n^+ in planetary nebulae using Eq. (67) gives values of the order of a few thousands. For these n^+ the ionization equation gives $\frac{n^+}{n_1} \approx 10^3$. Hence on the average there are a few neutral hydrogen atoms in 1 cm^3.

Knowledge of n^+ enables us to determine the mass of the emitting part of the nebula,

$$M = m_{\rm H} n^+ V, \tag{68}$$

where $m_{\rm H}$ is the mass of hydrogen atom. Inserting (67) in (68), we find

$$M = m_{\rm H} \sqrt{\frac{E_{k2} V}{z_k A_{k2} h\nu_{2k}}}. \tag{69}$$

The energy emitted in a given Balmer line is a certain fraction λ_{2k} of the visible luminosity of the nebula L, i.e., $E_{k2} = \lambda_{2k} L$. Eq. (69) can therefore be written in the form

$$M = C \sqrt{LV}, \tag{70}$$

where

$$C = m_{\rm H} \sqrt{\frac{\lambda_{2k}}{z_k A_{k2} h\nu_{2k}}}. \tag{71}$$

Eq. (70) was first derived by Ambartsumyan /14/. Application of this equation gives for the masses of planetary nebulae 0.01 $M_\odot$. Note that this

is the mass of the emitting part of the nebula. It corresponds to the total mass of the nebula if its optical thickness beyond the limit of the Lyman series is less than unity.

Vorontsov—Vel'yaminov /20/ determined the mass of NGC 6572 from its H_γ emission. He first had to determine the concentration of atoms as a function of the distance from the nucleus. His result was $M = 0.007\,M_\odot$.

The masses of diffuse gaseous nebulae lie between very wide limits, from a fraction of the solar mass to a few thousand solar masses. As an example, we would like to mention Guk's work /21/, who made photometry of the Omega Nebula NGC 6618 and determined its mass from Eq. (70). The mass of the brightest part of the nebula was found to be $30\,M_\odot$, and the mass of the entire nebula $515\,M_\odot$.

Once the number of hydrogen atoms in the nebula has been found we can determine the number of other atoms as well. This is accomplished by measuring the intensities of the lines of various atoms relative to the intensity of the Balmer lines in nebular spectra.

The intensity of a line produced by photoionization and recombination is expressed by Eq. (49). The parameter z_k entering this equation should be obtained from the stationarity equations for various levels of the atom, analogous to Eqs. (47). Calculation of z_k requires knowledge of the recombination probabilities and Einstein's coefficients of spontaneous transitions.

An example of lines of recombination origin is provided by the lines of neutral and ionized helium. The intensity of these lines relative to the Balmer line intensities indicates that the number of helium atoms in nebulae is on the average 1/10 of the number of hydrogen atoms /22, 9, 23/.

To find the intensities of lines associated with inelastic electron-atom collisions we require the effective cross sections for these collisions and Einstein's spontaneous-transition coefficients. In this case the number of excited atoms is determined by solving equations analogous to Eqs. (60) for O^{2+}.

As an example note that the ratio of the number of O^{2+} ions to the number of hydrogen atoms from the intensity ratio of $N_1 + N_2$ and H_β lines is found to be of the order of 10^{-4}.

A highly important fact is that nebular spectra show O II lines of recombination origin. The presence of these lines provides an independent estimate of the number of O^{2+} ions using Eq. (49). The two estimates show surprisingly good fit for different nebulae.

The most comprehensive treatment of the chemical composition of planetary nebulae was carried out by Aller and Menzel /22/. Their results are listed in Table 6.

TABLE 6

Element	Planetary nebula	Sun	τSco
Hydrogen	1000	1000	1000
Helium	100	222	175
Carbon	0.6	0.04	0.17
Nitrogen	0.2	0.12	0.3
Oxygen	0.25	0.37	1.0
Fluorine	0.0001		
Neon	0.01		0.1
Sulfur	0.036	0.037	
Chlorine	0.002		
Argon	0.0015		

For comparison the table also lists the chemical composition of the Sun and the star τSco obtained by an entirely different method—from absorption line intensities.

Recently Aller /23/ determined the chemical compsition of the planetary nebula NGC 7027, using Seatons's effective collision cross sections (see § 99). Allers results (logartithm of the relative number of atoms) are listed in Table 7, together with the chemical composition of a "normal star."

TABLE 7

Element	NGC 7027	Star	Element	NGC 7027	Star
H	13.4	13.27	S	9.21	8.52
He	12.0	12.0	Cl	8.39	8.3
N	9.56	9.49	A	8.22	9.0
O	10.00	10.00	K	6.2	6.39
F	7.4	7.7	Ca	6.6	7.72
Ne	9.48	10.04			

We see from the table that the differences in the chemical composition of nebulae and stellar atmospheres are not large.

§102. Radiation pressure in nebulae. As we have already noted in §96, the degradation of L_ϵ radiation in the nebula terminates with emission of L_α quanta. The optical thickness of the nebula in the Lyman lines being exceedingly large, these quanta diffuse through the nebula very slowly. The density of the L_α radiation in the nebula is therefore high.

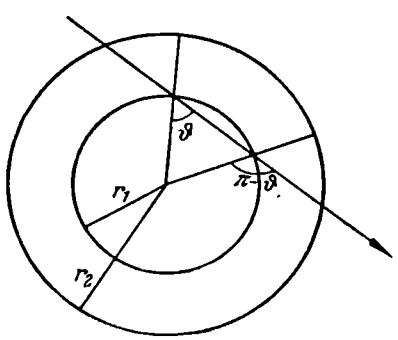


FIGURE 147

The problem of diffusion of L_α radiation in a nebula was first considered by Ambartsumyan /24/. He also noted the great importance of the L_α radiation pressure in the dynamics of nebulae.

Let us assess the radiation pressure forces at the outer boundary of a planetary nebula, which is assumed to be spherically symmetric (Figure 147).

When treating on radiation pressure in nebulae, we should distinguish between two cases: 1) the nebula is stationary, 2) the nebula expands with a velocity which is greater than the mean thermal velocity of atoms.

In case 1, the L_α quanta emerging through the inner surface of the nebula are absorbed in the opposite part of the nebula. Clearly, the intensity of radiation emerging at an angle ϑ to the normal through this surface is equal to the intensity of the incoming radiation incident at an angle $\pi-\vartheta$ to the normal; in other words the total flux of the L_α radiation at the inner surface of the nebula is zero. Hence, the L_α radiation pressure in this case does not vanish on the outer surface of the nebula only (where it is directed outward).

In case 2, the L_α quanta emerging through the inner surface are no longer absorbed by the opposite part of the nebula on account of the Doppler

shift. The L_α radiation pressure in this case does not vanish on either surface (on the outer surface the radiation pressure is directed outward, and on the inner surface it is directed inward, to the star).

Suppose that we are dealing with the second of the two cases above, in a nebula whose optical thickness beyond the limit of the Lyman series is much greater than unity. Then the majority of L_α quanta leave the nebula through its inner surface. We have noted before that the degradation of each L_c quantum in the nebula produces one L_α quantum. Therefore in this case the flux of L_α quanta at the inner surface of the nebula is equal (with opposite sign) to the flux of L_c quanta reaching the nebula from the star.

Consider a unit volume on the inner surface of the nebula. The radiation pressure acting on this volume is

$$R=\frac{n_1 k_{12}}{c} H_{12} h\nu_{12}, \tag{72}$$

where n_1 is the number of ground-state hydrogen atoms, k_{12} is the average absorption coefficient in the L_α line per atom, $h\nu_{12}$ is the energy of a L_α quantum, H_{12} is the flux of L_α quanta.

Since the L_α flux is equal to the L_c flux reaching the nebula from the star, we have

$$H_{12}=\left(\frac{r_*}{r_1}\right)^2 \frac{2\pi}{c^2} \int_{\nu_0}^{\infty} \frac{\nu^2\, d\nu}{e^{\frac{h\nu}{kT_*}}-1}, \tag{73}$$

where ν_0 is the hydrogen ionization frequency. Inserting (73) in (72) we get

$$R=\frac{2\pi n_1 k_{12} h\nu_{12}}{c^3}\left(\frac{r_*}{r_1}\right)^2 \int_{\nu_0}^{\infty} \frac{\nu^2\, d\nu}{e^{\frac{h\nu}{kT_*}}-1}. \tag{74}$$

Note that in calculating the radiation pressure on unit volume, we can ignore the L_c radiation, since the absorption coefficient in the L_α line is approximately 10^4 as great as the absorption coefficient beyond the Lyman limit.

Let us find the ratio of the radiation pressure force to the attraction of the central star. Since the unit volume contains ionized, as well as neutral, atoms, the attraction is given by

$$G=\left(\frac{r_*}{r_1}\right)^2 g_* m_{\mathrm{H}} (n_1+n^+), \tag{75}$$

where g_* is the gravitational acceleration on the star's surface.

For the ratio $\frac{R}{G}$ we get

$$\frac{R}{G}=\frac{2\pi k_{12} h\nu_{12}}{c^3 m_{\mathrm{H}} g_* \left(1+\frac{n^+}{n_1}\right)} \int_{\nu_0}^{\infty} \frac{\nu^2 d\nu}{e^{\frac{h\nu}{kT_*}}-1}. \tag{76}$$

Putting $T = 40{,}000°$, $\frac{n^+}{n_1} = 5{,}000$, we find

$$\frac{R}{G}=\frac{10^0}{g_*}. \tag{77}$$

g_* can hardly reach values of the order of 10^9 cm/sec^2 (the gravitational acceleration on the surface of the Sun, for example, is $3 \cdot 10^4$ cm/sec^2). We therefore conclude that the L_α radiation pressure in nebulae is clearly predominant compared to the attraction of the nucleus.

The radiation pressure should greatly slow down the parts of the expanding nebula closest to the star. Using Eq. (74), we can estimate this effect. The deceleration is found to be 3 km/sec per century.

The above calculations, however, are very crude, since they ignore the peculiar features of the diffusion of L_α radiation in nebulae.

We should remember that there is a definite velocity gradient in nebulae. If at some time all the strata of the nebula expanded at a constant velocity, the differential radiation pressure would eventually set up a range of expansion velocities. However, given different expansion velocities, the photons may diffuse through the central parts of the nebula without absorption owing to the Doppler effect. The L_α radiation density is thus lowered and the resulting L_α radiation pressure is less than that in uniformly expanding nebulae.

The diffusion of L_α quanta in a nebula with differential expansion velocities was first considered by Zanstra /25/. He assumed, however, that the frequency of the quantum did not change in diffusion. Later this problem was solved by the author /26/ assuming complete redistribution of radiation inside the line. It was shown that even a slight velocity gradient produces a very pronounced reduction in the density of the L_α radiation and the corresponding radiation pressure. In particular, the ratio of the radiation pressure on the inner surface in a nebula with differential expansion velocities to that in a uniformly expanding nebula was found to be

$$\sqrt{\frac{qu}{2\frac{dv}{d\tau}}},$$

where q is the ratio of the absorption coefficient beyond the Lyman limit to the L_α absorption coefficient, u is the average thermal velocity of atoms, $\frac{dv}{d\tau}$ is the velocity gradient in the nebula (τ being the optical thickness at the Lyman continuum frequencies). If u and $\frac{dv}{d\tau}$ are of the same order of magnitude, this ratio is of the order of 0.01.

Redistribution of radiation over the frequencies inside the line should clearly affect the radiation pressure values in a uniformly expanding nebula (or a stationary nebula) too. This problem also was considered by Zanstra /27/. Taking the absorption coefficient inside the line in the form

$$k_\nu = k_0 e^{-\left(\frac{\nu-\nu_0}{\Delta\nu}\right)^2},$$

he found that in a nebula with optical thickness of the order of unity in the Lyman continuum, the radiation pressure in case of complete redistribution over frequencies was about 1/300 of the radiation pressure without redistribution.

In a recent study /28/ the author investigated the diffusion of resonance radiation in a stationary nebula. Three cases of diffusion were treated: a) with unchanged radiation frequency, b) with complete redistribution

over frequencies, c) with actual redistribution over frequencies. In case b the emission coefficient ε_ν was taken to be proportional to the absorption coefficient k_ν. For case c we found the probability that the radiation absorbed by unit volume in the form of quanta of frequency ν_1 was emitted as quanta of frequency ν_2. The natural width of the atomic energy levels and the Doppler shift associated with the thermal motion of the atoms were taken into considerations. For each case the integral equation of diffusion was solved. Cases b and c were found to be close to each other, though greatly differing from case a. This in particular partly justifies the assumption of complete frequency redistribution, used in earlier studies /26,27/.

These results indicate that the L_α radiation pressure in nebulae is actually much less significant than suggested by Ambartsumyan's /24/ and Zanstra's /25/ findings. Its significance is definitely low in the ionized part of the nebula (i.e., where $\frac{n^+}{n_1} \gg 1$ for hydrogen atoms). The contribution from radiation pressure in unionized parts of the nebula may nevertheless be quite significant due to the increase in the number of absorbing atoms. This point was first emphasized by Gurzadyan in his work on the dynamics of nebulae /29/.

Gurzadyan studied in some detail a subgroup of planetary nebulae featuring a double envelope. These double-envelope nebulae have a number of remarkable properties: the mass of the outer envelope is invariably much less than the mass of the inner envelope, the outer envelope is immediately adjacent to the inner envelope, both envelopes are not greatly different from each other, etc. Examination of these properties leads to the conslusion that double-envelope nebulae definitely did not form as a result of successive ejection of matter from the star: they actually developed from a normal nebula at a certain stage of its evolution. Gurzadyan suggests that the second envelope is produced by the action of the L_α radiation pressure on the unionized component of the nebula in the period when its optical thickness in the Lyman continuum becomes of the order of unity. Gurzadyan's calculations bear out this hypothesis.

The L_α radiation pressure is thus capable of producing noticeable relative motion in different parts of the nebula. It may even somewhat accelerate the expansion of the entire nebula. If the nebula completely absorbs the stellar radiation beyond the Lyman limit, the velocity of the envelope produced by this radiation pressure is expressed by

$$v^2 - v_0^2 = \frac{2L}{Mc}(r - r_0), \tag{78}$$

where L is the luminosity of the star in the Lyman continuum, M is the mass of the nebula. Taking $L = 10^{36}$ erg/sec, $M = 0.001\, M_\odot$, $r - r_0 = 10{,}000$ a.u., $v_0 = 10$ km/sec, we obtain $v = 25$ km/sec. Observations indeed seem to indicate that the expansion of planetary nebulae is accelerated. This follows from a statistical comparison of the observed expansion velocities of nebulae with their dilution coefficients. These results are possibly associated with the effect of radiation pressure in the Lyman continuum (see Minin /30/).

Further note that the expansion of the nebula is affected by gas, as well as radiation, pressure.

§103. The continuous nebular spectrum. We have already noted that the spectra of gaseous nebula consist of emission lines on a continuous background. This continuous background is largely associated with recombinations and free-free transitions of electrons in the field of ions. The main contribution to this emission is from hydrogen, which is the most abundant element in nebulae.

As we know (Vol. I, §9), the quantity of energy emitted in 1 cm^3 in 1 sec at frequencies between ν and $\nu+d\nu$ when an electron is captured on the i-th level of a hydrogen atom is

$$n_e n^* \frac{2^9\pi^5}{(6\pi)^{\frac{3}{2}}} \frac{e^{10}}{m^2c^3h^2} \left(\frac{m}{kT_e}\right)^{\frac{3}{2}} \frac{1}{i^3} e^{\frac{\chi_i - h\nu}{kT_e}} d\nu \qquad (h\nu > \chi_i),$$

and the quantity of energy emitted by free-free transitions in 1 cm^3 in 1 sec in the same frequency range is

$$n_e n^* \frac{2^7\pi^3}{(6\pi)^{\frac{3}{2}}} \frac{e^6}{m^2c^3} \left(\frac{m}{kT_e}\right)^{\frac{1}{2}} e^{-\frac{h\nu}{kT_e}} d\nu.$$

In these expressions χ_i is the energy of ionization from the i-th state, T_e is the temperature of the electron gas, n_e and n^* is the number of free electrons and protons in 1 cm^3, respectively.

The total quantity of energy emitted in a unit frequency interval in 1 cm^3 in 1 sec is given by

$$n_e n^* \frac{2^7\pi^3}{(6\pi)^{\frac{3}{2}}} \frac{e^6}{m^2c^3} \left(\frac{m}{kT_e}\right)^{\frac{1}{2}} \left(1 + 2\frac{\chi_j}{kT_e} \sum_{i=j}^{\infty} \frac{1}{i^3} e^{\frac{\chi_i}{kT_e}}\right) e^{-\frac{h\nu}{kT_e}}. \qquad (79)$$

Here $j = 2$ at frequencies between the Balmer and the Lyman limit, $j = 3$ between the Paschen and the Balmer limit, etc.

If the continuous emission of the nebula follows Eq. (79), comparison of theoretical and observed spectrophotometric temperatures and the intensity jump near the Balmer limit will give the electron temperature of the nebula.

The spectrophotometric temperature T_s is related to the electron temperature T_e by the equality

$$-\frac{h}{kT_e} = \frac{3}{\nu} - \frac{h}{kT_s} \frac{1}{1 - e^{-\frac{h\nu}{kT_s}}}. \qquad (80)$$

This relation is obtained when the expression for the radiation flux

$$H_\nu \sim e^{-\frac{h\nu}{kT_e}}, \qquad (81)$$

which follows from (79), is substituted in Eq. (80) of Ch. XIV (the definition of spectrophotometric temperature).

The Balmer jump in this case is given by

$$D = \lg \frac{1 + 2\frac{\chi_1}{kT_e}\sum_{i=3}^{\infty}\frac{1}{i^3}e^{\frac{\chi_i}{kT_e}}}{1 + 2\frac{\chi_1}{kT_e}\sum_{i=2}^{\infty}\frac{1}{i^3}e^{\frac{\chi_i}{kT_e}}}. \qquad (82)$$

Table 8 lists the Balmer jump D for various electron temperatures T_e, at various values of the ratio C/Ba_C — the intensity of some other continuum (if present) to the intensity of the Balmer continuum. The table is borrowed from Greenstein /31/

TABLE 8

T_e	C/Ba_C				T_e	C/Ba_C			
	0	0.1	0.2	0.3		0	0.1	0.2	0.3
5000°	2.34	1.02	0.77	0.63	20 000°	0.72	0.58	0.49	0.42
7 500	1.68	0.96	0 73	0.61	25 000	0.60	0.49	0.42	0.37
10 000	1.31	0.87	0.68	0.57	30 000	0.51	0.41	0.36	0.32
15 000	0.93	0.70	0.57	0.49	40 000	0.38	0.33	0.29	0.26

The continuous spectra of planetary nebulae were studied by Page /32, 33/. He measured the energy distribution in nebular spectra beyond the Balmer limit and found that, if corrected for interstellar absorption, it on the whole fits Eq. (80). Page also discovered a continuous spectrum of approximately constant intensity between λ3600 Å and λ4800 Å, which could not be explained by the recombination mechanism.

Barbier /34/ and Greenstein /31/ studied the continuous spectrum of the Orion Nebula. For the central part of the nebula Greenstein found $T_s = 12{,}000°$ on either side of the Balmer limit. According to Barbier's measurements between λ4190 Å and λ4685 Å, $T_s = 6{,}950°$ (too low in Greenstein's opinion), and on the violet side of the Balmer limit $T_s = 10{,}300°$. The electron temperature of the nebula according to Eq. (80) is $T_e = 65{,}000°$ for $T_s = 12{,}000°$ and $T_e = 33{,}000°$ for $T_s = 10{,}300°$. These values of T_e are highly different from the electron temperatures obtained by other methods (see §100).

According to Greenstein, the Balmer jump in the spectrum of the Orion Nebula is $D = -0.64$ ($D = -0.56$ according to Barbier). For this D, the table gives $T_e = 22{,}000°$. This electron temperature is again much higher than the values obtained, say, from the $(N_1 + N_2)/\lambda$ 4363 Å intensity ratio. To avoid this divergence, we have to assume that the continuous nebular spectrum is produced by the joint action of recombinations together with some other mechanism. If, for example, we take $C/Ba_C = 0.2$, we get $T_e = 12{,}000°$.

It follows from the preceding that the origin of the continuous spectrum of gaseous nebulae cannot be attributed to recombinations and free-free transitions alone. In some diffuse nebulae, dust particles scattering the stellar light are an additional source of glow. Planetary nebulae, however, are apparently free from dust.

Recently Kipper /35/, and also Spitzer and Greenstein /36/ described a new mechanism responsible for the continuous emission of gaseous nebula. It involves the emission of two quanta by a hydrogen atom jumping from the $2s$ state to $1s$.

Let the frequencies of the two quanta emitted in the $2s \rightarrow 1s$ transition be $y\nu_{12}$ and $(1-y)\nu_{12}$, where ν_{12} is the L_α frequency. Let $A(y)dy$ be the transition probability associated with the emission of a quantum at frequencies between $\nu_{12}y$ and $\nu_{12}(y+dy)$. Writing

$$A(y) = \frac{9\alpha^6\nu_0}{2^{10}}\psi(y), \tag{83}$$

where ν_0 is the hydrogen ionization frequency and $\alpha = \frac{2\pi e^2}{hc}$ is the fine structure constant, the above authors calculated the function $\psi(y)$(Table 9). The energy emitted in unit frequency interval is proportional to $h\nu A(y)$ or $y\psi(y)$. The values of $y\psi(y)$ are also given in Table 9. Einstein's transition coefficient from state $2s$ to state $1s$ is

$$A_{2s,\,1s} = \frac{1}{2}\int_0^1 A(y)\,dy = 8.227\ \text{sec}^{-1}. \tag{84}$$

TABLE 9

y	λ (Å)	$\psi(y)$	$y\psi(y)$
0.00		0	0
0.05	24313	1.725	0.0863
0.10	12157	2.783	0.2783
0.15	8105	3.481	0.5222
0.20	6078	3.961	0.7922
0.25	4862	4.306	1.077
0.30	4052	4.546	1.363
0.35	3473	4.711	1.649
0.40	3039	4.824	1.929
0.45	2702	4.889	2.200
0.50	2403	4.907	2.454

In gaseous nebulae the hydrogen atom is excited to the $2s$ state following recombinations and cascade transitions of electrons from higher to lower states. The fraction of electrons passing through the state $2s$ on their way down relative to the total number of electrons captured on second and higher levels can be found without difficulty. Calculations give for this fraction $X = 0.32$.

The values of X and $A(y)$ can be used to determine the quantity of energy emitted in $2s \rightarrow 1s$ transitions by unit volume in 1 sec at various frequencies. This energy is found to be comparable with the energy emitted in recombinations and free-free transitions.

The mechanism of two-photon emission largely accounts for the continuous nebular spectrum between λ3600 Å and λ4800 Å, discovered by Page. There is, however, no perfect fit between theory and observations, possibly due to inaccuracy of the latter. Two-photon emission can be taken into consideration also in the theoretical determination of spectrophotometric temperatures and Balmer jumps.

Detailed comparison of theory with experiment should take account of the fact that a hydrogen atom in the $2s$ state does not necessarily drop to the $1s$ state with emission of two quanta in the continuous spectrum. De-excitation of the $2s$ state may also be caused by collisions. If only collisions with free electrons are taken into consideration, we obtain

$$X = \frac{0.32}{1+8.2\cdot 10^{-6} n_e}, \tag{85}$$

where n_e is the number of free electrons in 1 cm^3. In this case, however, collisions with protons are of greater significance than collisions with electrons, and we get

$$X = \frac{0.32}{1+0.6\cdot 10^{-4} n_e}, \tag{86}$$

where the concentration of protons is taken equal to the concentration of free electrons.

Seaton /37/ applied the above equations to calculate the Balmer jump for some nebulae. His results are listed in Table 10. The last column gives the theoretical values of the Balmer jump calculated for the following three cases: 1) using recombinations and free-free transitions, 2) introducing two-photon emission with $X = 0.32$, 3) introducing two-photon emission with X from Eq. (86).

TABLE 10

Nebula	$10^{-4}T_e$	$10^{-4}n_e$	$-D_{obs}$	$-D_{theo.}$
NGC 6543	1.0	3	0.98	1.26 0.70 0.95
6572	1.3	5	0.79	1.00 0.59 0.84
6826	1.1	3	0.61	1.15 0.66 0.89
7009	1.4	3	0.82	0.90 0.56 0.73
7662	1.3	5	0.81	0.80 0.59 0.79
IC 418	1.9	0.8	0.48	0.69 0.45 0.50
Average			0.75	0.98 0.59 0.78

We see from the table that two-photon emission markedly affects the value of the Balmer jump. There is furthermore good fit between theory and observations for X from Eq. (86).

The Crab Nebula — a remnant of the supernova explosion of 1054 — is an object of particular interest. Its continuous spectrum is very strong; the energy emitted in the line spectrum is but a few percent of the energy emitted in the continuous spectrum. Minkowski /38/ made an attempt to interpret the continuous spectrum of the Crab Nebula as an outcome of recombinations and free-free transitions. He obtained in this case 15 $M_\odot$ for the mass of the nebula and a kinetic temperature of a few hundred thousand degrees. The central star was found to possess most peculiar properties too: its temperature was found to be 500,000°, its radius 0.02 $R_\odot$.

New findings, however, led to a radical revision of our interpretation of the continuous spectrum of the Crab Nebula. In 1949 the Crab Nebula was found to be a powerful radio source. Somewhat later, a new radiation

mechanism was proposed — the bremsstrahlung of relativistic electrons in magnetic fields. This led Shklovskii /39/ to the hypothesis that the radio emission of the Crab Nebula, as well as its emission in the visible spectrum, is associated with the radiation of relativistic electrons (with energies reaching 10^{11} eV and higher). The origin of the relativistic electrons in the Crab Nebula is not clear at this stage.

Note that the unusual origin of the continuous spectrum of the Crab Nebula also follows from the results of Dombrovskii /40/, who measured the polarization of radiation from different parts of the nebula. He found that although the percentage polarization differed in different parts of the nebula (9—15%), the direction of polarization is mostly the same. In his opinion, this result can be understood if the continuous spectrum of the Crab Nebula is interpreted as the radiation of relativistic electrons.*

REFERENCES TO CHAPTER XVI

1. BOWEN, I. The origin of the nebular lines and the structure of the planetary nebulae.— ApJ 67:1. 1928.
2. SHAIN, G.A. and V.F. GAZE. Some results concerning luminous galactic nebulae.— Izvestiya, Crimean Astrophysical Observatory, 6-3. 1951; The second list of diffuse nebulae.— Izvestiya, Crimean Astrophysical Observatory, 7:93. 1951.
3. VORONTSOV-VEL'YAMINOV, B.A. The system of planetary nebulae.— Astron. Zh., 27:285. 1950.
4. ZANSTRA, H. Die Leuchtprozesse planetarisches Nebel und die Temperatur der Zentralsterne.— Zs. f. Astrophys., 2:1. 1931.
5. AMBARTSUMYAN, V.A. On the temperatures of the nuclei of planetary nebulae.— Tsirkulyar, Pulkovo Observatory, No.4. 1932.
6. CILLIE, G. The theoretical capture spectrum of hydrogen.— MN 96:771. 1936.
7. BAKER, J. and D. MENZEL. The Balmer decrement.— ApJ 88:52. 1938,
8. GOLDBERG, L. The statistical equilibrium of neutral helium.— ApJ 93:244. 1941.
9. NIKITIN, A.A. The abundance of helium in planetary nebulae.— Uchenye Zapiski, Leningrad University, No.136. 1950.
10. AMBARTSUMYAN, V.A. The excitation of the metastable states in the gaseous nebulae.— Tsirkulyar, Pulkovo Observatory, No.6. 1933.
11. GAUZIT, J. Sur les conditions d'emission des radiations interdites.— Ann. d'astrophys. 6:53. 1943.
12. NIKITIN, A.A. The mass of Orion Nebula.— Vestnik, Leningrad University, No.5. 1954.
13. SEATON, M.J. Electron excitation of forbidden lines occuring in gaseous nebulae.— Proc. Roy. Soc., 218A, N1134. 1953.
14. AMBARTSUMYAN, V.A. Teoreticheskaya astrofizika (Theoretical Astrophysics).— GONTI. 1939.
15. MENZEL, D., L. ALLER and M. HEBB. The electron temperatures of some typical planetary nebulae.— ApJ 93:230. 1941.
16. SEATON, M.J. Electron temperatures and electron densities in planetary nebulae.— MN 114:154. 1954.
17. ALLER, L.H. Thermal effects of collisional line excitation in gaseous nebulae.— ApJ 118:547. 1953.
18. SOBOLEV, V.V. Determination of the electron temperatures of planetary nebulae and improvement of the temperatures of their nuclei.— Uchenye Zapiski, Leningrad University, No.82. 1941.
19. SOBOLEV, V.V. Dvizhushchie obolochki zvezd (Moving Star Envelopes).— Leningrad University. 1947.
20. VORONTSOV-VEL'YAMINOV, B.A. Spatial structure and mass of the planetary nebula NGC 6572.— Zs. F. Astrophys., 12-247. 1936.
21. GUK, D.P. Determination of the mass of the gaseous diffuse nebulae Omega NGC 6618.— Astron. Zh., 28:253. 1951.
22. ALLER, L. and D. MENZEL. The chemical composition of the planetary nebulae.— ApJ 102:239. 1945.
23. ALLER, L. The composition of the planetary nebulae NGC 7027.— ApJ 120:401. 1954.
24. AMBARTSUMYAN, V.A. On the radiative equilibrium of a planetary nebula. — Tsirkulyar, Pulkovo Observatory, 13:31. 1933.

25. ZANSTRA, H. Radiation pressure in an expanding nebula.— MN 95:84. 1934.
26. SOBOLEV, V.V. Radiation pressure in an expanding nebula.— Astron. Zh., 22:143. 1944.
27. ZANSTRA, H. On scattering with redistribution and radiation pressure in a stationary nebula.— Bull. Astr. Inst. Netherland XI, N401. 1949.
28. SOBOLEV, V.V. Diffusion of radiation with redistribution over frequencies.— Vestnik, Leningrad University, No.5. 1955.
29. GURZADYAN, G.A. Voprosy dinamiki planetarnykh tumannostei (The Dynamics of Planetary Nebulae).— Erevan. 1954.
30. MININ, I.N. Ejection of matter from Wolf— Rayet stars.— In: Proc. Fourth Conf. on Problems of Cosmogony, 1955.
31. GREENSTEIN, J.L. Continuous emission in the Orion nebula.— ApJ 104:414. 1946.
32. PAGE, T. The continuous spectra of certain planetary nebulae.— MN 96:604. 1936.
33. PAGE, T. Continuous emission in the spectra of planetary nebulae.— ApJ 96:78. 1942.
34. BARBIER, D. Etude photometrique du spectre continu de la nebuleuse d'Orion.— Ann. d'Astrophys., 7:80. 1944.
35. KIPPER, A.Ya. Continuous spectrum of planetary nebulae.— In: "O razvitii sovetskoi nauki v Estonskoi SSR 1940—1950." 1950.
36. SPITZER, L. and J. GREENSTEIN. Continuous emission from planetary nebulae.— ApJ 114:407. 1951.
37. SEATON, M.J. Deactivation by collisions of the 2s metastable state of hydrogen in planetary nebulae.— Ann. astrophys., 17:269. 1954.
38. MINKOWSKI, R. Crab Nebula.— ApJ 96:199. 1942.
39. SHKLOVSKII, I.S. The origin of emission of the Crab Nebula.— Doklady AN SSSR, 90:983. 1953.
40. DOMBROVSKII, V.A. The origin of emission of the Crab Nebula.— Doklady AN SSSR, 94:1021. 1954.

V. V. SOBOLEV

Chapter XVII

STARS WITH BRIGHT SPECTRAL LINES

§104. Early-type stars with bright lines. After a discussion of gaseous nebulae, we now return to stellar atmospheres. In this chapter, however, we are no longer concerned with "normal" stars with absorption spectra, but rather with "extraordinary" stars with spectra showing bright emission lines. The results obtained for gaseous nebulae are of fundamental significance in understanding the processes which take place in the atmospheres of these stars.

We will first discuss early-type stars with bright lines, which include Wolf—Rayet, P Cygni, and Be stars. This group has been studied in much greater detail than other stars with bright spectral lines.

Wolf—Rayet stars are classified in spectral type O according to the degree of excitation and ionization of their atoms. Their spectra consist of a continuous background with superimposed wide emission bands of H, He I, He II, C III, N III and other atoms with extremely high ionization potentials. Some of the bright bands have violet satellites in the form of weak absorption lines. The line profiles in the spectra of Wolf—Rayet stars are shown in Figure 148, after Beals /1/.

Beals established that the spectra of Wolf—Rayet stars fall into two distinct sequences: the nitrogen sequence and the carbon sequence. The spectra of the nitrogen sequence show bands of nitrogen in various ionization stages, whereas carbon and oxygen bands are absent; the spectra of the second sequence carry bands of carbon and oxygen in various stages of ionization, but no nitrogen bands. The degree of excitation and ionization of the atoms in the two sequences is roughly the same. Wolf—Rayet spectra with both nitrogen and carbon bands were subsequently discovered. Nevertheless, some spectra are apparently characterized by stronger nitrogen lines, and others by stronger carbon bands.

Wolf—Rayet stars are often found to be components of spectroscopic binaries. Some of these are also eclipsing variables. The study of these binary systems provided a wealth of information on Wolf—Rayet stars. In particular, it was established that their masses are generally of the order of 10 solar masses.

The visual absolute magnitudes of Wolf—Rayet stars are about -3^m. These stars are among the brightest objects in the Galaxy. Wolf—Rayet spectra, however, are characteristic not only of these stars: novae even a few years after the eruption and some nuclei of planetary nebulae also show Wolf—Rayet spectra, although their visual absolute magnitudes are on the average $+5^m$.

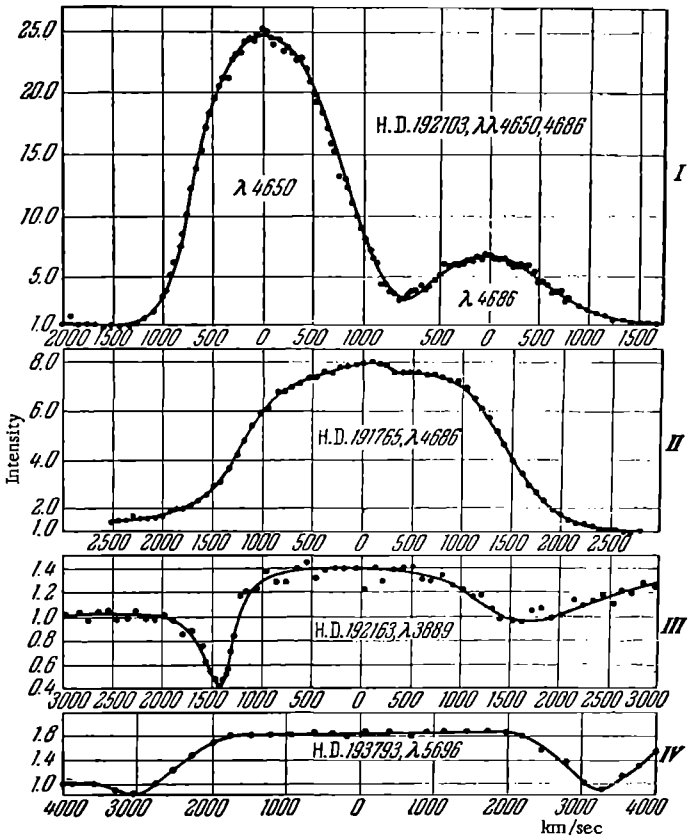


FIGURE 148

Wolf—Rayet stars adjoin P Cygni stars, which belong to spectral type B. The spectra of these stars, like Wolf—Rayet spectra, show bright symmetric lines with a violet absorption satellite (Figure 149a). The bright lines, however, are not as wide as in Wolf—Rayet spectra, whereas the absorption lines are conversely much stronger.

P Cygni stars are not the only B-type stars with emission lines. All the rest are generally designated as Be stars. The line profiles in Be spectra can be described as a superposition of an emission line on a wider and relatively shallow absorption line; in some spectra the emission line has a single peak, while in others it is split into two (Figure 149b, c). Be spectra undergo marked variations with time, e.g., the relative intensities of the bright line components change. Sometimes the bright lines disappear altogether, and the Be star is transformed into a normal B-type star. The spectral variations are accompanied by slight fluctuations in stellar magnitude.

It is generally accepted that gaseous matter is ejected from these extraordinary stars (continuous ejection from Wolf—Rayet and P Cygni stars, and irregular, intermittent ejection from Be stars). This produces

extended moving envelopes around the stars. The physical conditions in these envelopes are not unlike those in gaseous nebulae. In particular, they are characterized by a low dilution coefficient. In these stellar envelopes, like in nebulae, high-frequency quanta are therefore degraded to low-frequency quanta. Since the stars of these groups are extremely hot, i.e., emit large quantities of high-frequency energy, their visible emission is correspondingly very strong. This explains the origin of the observed emission lines in the spectra of these stars.

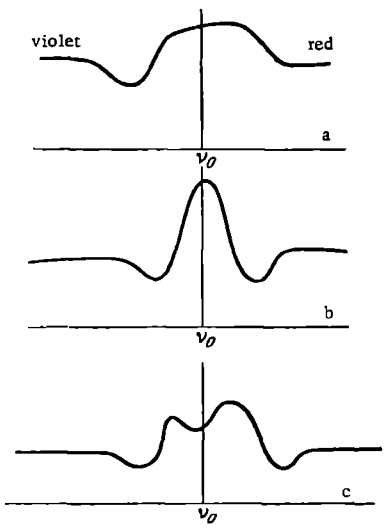


FIGURE 149

Detailed interpretation of the spectra of Wolf—Rayet, P Cygni, and Be stars (as well as of other nonstationary stars) requires a consistent theory of spectral lines in extended moving stellar envelopes. The elements of this theory are presented in the following sections.

§105. Emission line profiles. The profiles of emission lines originating in moving shells are determined assuming that the envelope velocity is much greater than the thermal velocity of atoms (this assumption is generally justified in practice). In this case the line profile is determined mainly by the motion of the envelope. The effect of other factors on the line profile can be ignored to first approximation.

Emission line profiles for the case of a transparent envelope can be calculated without difficulty. This case was treated by Beals /1/, Chandrasekhar /2/, Gerasimovich /3/, and others. The envelope was regarded as having spherical symmetry and the atoms were assumed to move in the radial direction only. The results gave the line profiles for different density and velocity distributions in the envelope. If the velocity of the atoms is independent of the distance from the star center, the line has a rectangular profile (i.e., the intensity is constant inside the line). If there is a velocity gradient in the envelope, the line profile is symmetric, with intensity falling

off on either side of the line center. If the partial screening of the envelope by the star is taken into account, the line is found to be asymmetric and slightly shifted toward the violet.

A determination of emission line profiles for opaque envelopes was carried out by the author /4/. Ambartsumyan et al. /5/ considered line profiles in moving envelopes, both transparent and opaque to the line radiation. An arbitrary velocity field was assumed in these studies. The approximate solution of this problem is as follows.

Consider a line associated with the $k \to i$ transition in a given atom. The absorption coefficient α_{ik} and the emission coefficient ε_{ik} are assumed constant for $\nu_{ik} - \frac{\Delta\nu_{ik}}{2} < \nu < \nu_{ik} + \frac{\Delta\nu_{ik}}{2}$ and zero elsewhere. Here ν_{ik} is the central line frequency, $\Delta\nu_{ik} = 2\frac{u}{c}\nu_{ik}$, where u is the mean thermal velocity of the atom, c is the velocity of light.

Consider an xyz frame with the origin at the center of the star and the z axis directed toward the observer. The velocity of atoms in the envelope will be designated $v(x, y, z)$ and its projection on the z axis $v_z(x, y, z)$. Let $v \gg u$.

Under the above assumptions on α_{ik} and ε_{ik}, radiation of frequency ν is emitted toward the observer only by the part of the envelope on the two sides of the surface of equal radial velocities defined by the equation

$$\nu = \nu_{ik} + \frac{\nu_{ik}}{c} v_z(x, y, z). \tag{1}$$

The line-of-sight half-thickness of the emitting region (i.e., the distance in the direction of the z axis from the surface (1)) corresponds to a frequency increment of $\frac{\Delta\nu_{ik}}{2}$. Writing z_1 and z_2 for the limit values of z and seeing that u is small compared to v, we get

$$\Delta\nu_{ik} = \frac{\nu_{ik}}{c}\left|\frac{\partial v_z}{\partial z}\right|(z_2 - z_1), \tag{2}$$

or

$$z_2 - z_1 = \frac{2u}{\left|\frac{\partial v_z}{\partial z}\right|}. \tag{3}$$

Let $I_{ik}(x, y, \nu)$ be the intensity of radiation from a point x, y of the disk at the frequency ν inside the line. Since the thickness of the layer emitting at the frequency ν (i.e., the difference $z_2 - z_1$) is relatively small (except at isolated points), α_{ik} and ε_{ik} may be regarded constant in the z direction across the layer; for these coefficients we may thus take their values on the surface (1). Thus,

$$I_{ik}(x, y, \nu) = \frac{\varepsilon_{ik}}{\alpha_{ik}}\left[1 - e^{-\alpha_{ik}(z_2 - z_1)}\right]. \tag{4}$$

The total energy emitted by the envelope at the frequency ν in a unit solid angle is

$$E_{ik}(\nu) = \int\int I_{ik}(x, y, \nu)\, dx\, dy. \tag{5}$$

Using (3) and (4), we obtain from (5)

$$E_{ik}(\nu)=\int\int \frac{\varepsilon_{ik}}{\alpha_{ik}}\left[1-e^{-\frac{2u}{\left|\frac{\partial v_z}{\partial z}\right|}\alpha_{ik}}\right]dx\,dy. \tag{6}$$

The integration is over the surface (1). Eq. (6) describes the emission line profile.

In general, the envelope can be divided into two parts: the part opaque to the radiation in the particular line and the part transparent to the line radiation. In the first region, $\frac{2u}{\left|\frac{\partial u_z}{\partial z}\right|}\alpha_{ik}$ is greater than unity and in the second it is less than unity. The integral (6) in the first region is

$$E'_{ik}(\nu)=\int\int \frac{\varepsilon_{ik}}{\alpha_{ik}}\,dx\,dy, \tag{7}$$

and in the second region

$$E''_{ik}(\nu)=\int\int \varepsilon_{ik}\frac{2u}{\left|\frac{\partial v_z}{\partial z}\right|}\,dx\,dy. \tag{8}$$

The coefficients α_{ik} and ε_{ik} entering the above expressions are related to the concentrations of the absorbing and the emitting atoms (n_i and n_k, respectively):

$$\varepsilon_{ik}=\frac{n_k A_{ki} h\nu_{ik}}{4\pi\Delta\nu_{ik}}, \tag{9}$$

$$\alpha_{ik}=\frac{n_i B_{ik} h\nu_{ik}}{\Delta\nu_{ik}c}\left(1-\frac{g_i}{g_k}\frac{n_k}{n_i}\right), \tag{10}$$

where A_{ki} and B_{ik} are Einstein's transition coefficients. Using the relation between A_{ki} and B_{ik}, we get

$$\frac{\varepsilon_{ik}}{\alpha_{ik}}=\frac{2h\nu_{ik}^3}{c^2}\frac{1}{\frac{g_k}{g_i}\frac{n_i}{n_k}-1}. \tag{11}$$

Eq. (11) quite properly reduces to Planck's equation when $\frac{n_k}{n_i}$ is expressed from the Boltzmann distribution.

Thus, calculation of emission line profiles requires knowledge of the velocity distribution in the envelope and of the distribution of the absorbing and emitting atoms. In the next section we will show how n_i and n_k can be found. This will essentially complete the determination of the emission line profiles.

As an example of the application of Eq. (6), let us find the emission line profiles originating in a uniformly expanding envelope ($v=$ const). Let r be the distance of a given volume element from the star's center and ϑ the angle between the direction of motion of the atoms and the direction to the observer. We have in this case $v_z=v\cos\vartheta$, $\frac{\partial v_z}{\partial z}=\frac{v}{r}\sin^2\vartheta$, and the surface of equal radial velocities corresponding to the frequency ν is given by the equation

$$\nu=\nu_{ik}+\frac{\nu_{ik}}{c}v\cos\vartheta. \tag{12}$$

From Eq. (6) we obtain for a transparent envelope

$$E_{ik}'(\nu) = 4\pi \frac{u}{v} \int \varepsilon_{ik} r^2 \, dr, \tag{13}$$

and for an opaque envelope

$$E_{ik}''(\nu) = 2\pi \sin^2 \vartheta \int \frac{\varepsilon_{ik}}{\alpha_{ik}} r \, dr,$$

or using (12)

$$E_{ik}''(\nu) = 2\pi \left[1 - \left(\frac{c}{v} \frac{\nu - \nu_0}{\nu_0} \right)^2 \right] \int \frac{\varepsilon_{ik}}{\alpha_{ik}} r \, dr. \tag{14}$$

The transparent envelope thus gives a rectangular line profile (as mentioned in the preceding), whereas the opaque envelope gives an emission line of parabolic profile. The widths of the two lines correspond to double the expansion velocity of the envelope.

The emission line profiles in an envelope where v is a function of r are obtained by the same method. For transparent envelopes the line profiles are reminiscent of those in opaque envelopes with $v = \text{const}$, so that before applying the line profile to determine the velocity distribution in the envelope, we must establish whether the envelope is transparent or opaque to the line radiation.

Using Eq. (6) we can also determine the profiles of emission lines originating in an envelope ejected by a rotating star. This problem is related to the interpretation of the spectra of Be stars.

§106. Emission line intensities. In §98 we determined the intensities of emission lines of gaseous nebulae. The problem was reduced to the solution of the system of linear algebraic equations

$$n_i \sum_{k=2}^{i-1} A_{ik} = \sum_{k=i+1}^{\infty} n_k A_{ki} + n_e n^+ C_i(T_e), \tag{15}$$

which provide an expression of the stationarity conditions for each excited level. Here n_i, n_e, n^+ are the concentrations of atoms in the i-th state, free electrons, and ions, respectively; $n_e n^+ C_i(T_e)$ is the number of electrons captured on the i-th level in 1 cm^3 in 1 sec. In the derivation of these equations the nebula was assumed transparent to the line emission of the subordinate series. This assumption is acceptable, since the degree of excitation of atoms in nebulae is very low ($\frac{n_i}{n_1}$ is proportional to the dilution coefficient W).

The emission lines in the spectra of Wolf—Rayet, P Cygni and Be stars are produced by fundamentally the same mechanism as the emission line in nebular spectra, namely by photoionization and recombination. The degree of excitation of atoms in the stellar envelopes, however, is not as low as in nebulae, so that the envelope may prove to be opaque to radiation in the lines of subordinate series. For this reason, Eqs. (15) do not always apply to the determination of line intensities in stellar spectra. In general, a new set of equations should be drawn up, which allow for the opacity of the envelope to line radiation; in other words, in addition to the stationarity equations for each level, we should introduce the equation of transfer in each line.

The motion of the envelope complicates the problem even further. However, if the velocity gradient in the envelope is fairly high (as in the case in the stars under discussion), the problem is substantially simplified. In an envelope with differential expansion, the emission line quanta diffuse from the central parts (due to the Doppler effect), as well as from the periphery. Roughly speaking, the velocity gradient improves the transparency of the envelope to line emission. The determination of emission line intensities therefore again reduces to a system of algebraic equations (though not linear as for nebulae) /6/.

To draw up the required equations, we should consider the following processes: 1) ionization of atoms from each state by stellar radiation; 2) recombinations to each state; 3) spontaneous transitions from upper to lower states; 4) transitions from lower to higher states induced by the line emission of the envelope.

The number of spontaneous $k \to i$ transitions in 1 cm^3 in 1 sec is $n_k A_{ki}$. If the envelope is transparent to the radiation in the particular line, the quanta emitted in this transition escape unobstructed from the envelope, and no reverse transitions are possible. If the envelope is opaque to this radiation and its velocity is constant, almost all the line quanta are absorbed in the envelope and the number of $k \to i$ transitions is almost exactly equal to the number of $i \to k$ transitions. In the presence of a velocity gradient, however, a certain proportion of the line quanta are lost owing to the Doppler effect. This fraction will be designated β_{ik}. The number of $k \to i$ transitions is then greater by $n_k A_{ki} \beta_{ik}$ than the number of reverse transitions.

Since the number of transitions from the i-th state to all the other states should equal the number of transitions to the i-th state, we have

$$n_i \Big(\sum_{k=1}^{i-1} A_{ik} \beta_{ki} + B_{ic} \varrho_{ic} \Big) = \sum_{k=i+1}^{\infty} n_k A_{ki} \beta_{ik} + n_e n^+ C_i (T_e), \tag{16}$$

where $n_i B_{ic} \varrho_{ic}$ is the number of ionizations from the i-th state. ϱ_{ic} are assumed known,

$$\varrho_{ic} = W \varrho_{ic}^*, \tag{17}$$

where ϱ_{ic}^* is the radiation density beyond the limit of the i-th series in the steller atmosphere and W is the dilution coefficient.

In the determination of β_{ik} we assume, as before, that α_{ik} and ε_{ik} both have constant finite values inside the frequency interval $\Delta\nu_{ik} = 2\frac{u}{c}\nu_{ik}$ and are zero elsewhere. We moreover assume that the part of the envelope where the relevant line radiation is absorbed is relatively small (on account of the large velocity gradient), so that the density of matter and the velocity gradient in this region can be assumed constant.

Let us consider the outgoing radiation of frequency ν_{ik} from some volume element in the direction inside a solid angle $d\omega$. On traversing the distance from s to $s + ds$, the radiation flux loses through absorption the following fraction of the emitted quanta:

$$e^{-\alpha_{ik} s} \Big(1 - \frac{|\nu'_{ik} - \nu_{ik}|}{\Delta\nu_{ik}} \Big) \alpha_{ik}\, ds, \tag{18}$$

where the factor $e^{-\alpha_{ik}s}$ accounts for absorption between 0 and s, and the factor $1 - \frac{|\nu'_{ik}-\nu_{ik}|}{\Delta\nu_{ik}}$ represents the change in frequency due to the Doppler effect. Here

$$\nu'_{ik} - \nu_{ik} = \frac{\nu_{ik}}{c}\frac{\partial v_s}{\partial s}s. \tag{19}$$

The fraction of quanta absorbed across the entire envelope is thus

$$\int_0^{s_1} e^{-\alpha_{ik}s}\left[1 - \frac{1}{2u}\left|\frac{\partial v_s}{\partial s}\right| s\right]\alpha_{ik}\,ds, \tag{20}$$

where s_1 is given by

$$\frac{1}{2u}\left|\frac{\partial v_s}{\partial s}\right| s_1 = 1. \tag{21}$$

Multiplying Eq. (20) by $\frac{d\omega}{4\pi}$ and integrating over the solid angles, we obtain the proportion of quanta absorbed in the envelope relative to the total number of quanta emitted by a given volume. In our notation, this fraction is $1-\beta_{ik}$. For β_{ik} we thus get

$$\beta_{ik} = \int\left(1 - e^{-\frac{2u}{\left|\frac{\partial v_s}{\partial s}\right|}\alpha_{ik}}\right)\frac{1}{2u\alpha_{ik}}\left|\frac{\partial v_s}{\partial s}\right|\frac{\partial\omega}{4\pi}. \tag{22}$$

If the envelope is opaque in all directions at a given point, we have

$$\beta_{ik} = \frac{1}{2u\alpha_{ik}}\left|\overline{\frac{\partial v_s}{\partial s}}\right|. \tag{23}$$

where $\left|\overline{\frac{\partial v_s}{\partial s}}\right|$ is the value of $\left|\frac{\partial u_s}{\partial s}\right|$ averaged over all the directions. If the envelope is transparent in all directions at the given point (i.e., $\frac{2u}{\left|\frac{\partial u_s}{\partial s}\right|}\alpha_{ik} \ll 1$), we properly find $\beta_{ik} = 1$.

We have thus obtained Eqs. (16) for the unknowns n_i, where β_{ik} are determined by relations (22). The absorption coefficients α_{ik} entering these relations are given by Eq. (10).

If Eqs. (16) have been solved for different parts of the envelope, we can find the total quantity of energy emitted by the envelope in any spectral line. This is done using the expression

$$E_{ki} = A_{ki}h\nu_{ik}\int n_k\beta_{ik}\,dV, \tag{24}$$

where the integration is carried out over the entire volume of the envelope. For a transparent envelope $\beta_{ik} = 1$ and Eq. (24) reduces to Eq. (48) of the previous chapter.

Eqs. (16) can be solved numerically. To this end we require four parameters: the star's temperature T_* (which determines ϱ^*_{ic}), the electron temperature T_e (which determines C_i), the dilution coefficients W, and the coefficient β_{12}. Table 1 lists the values of the Balmer decrement obtained

for $T_* = 20{,}000°$, $T_e = 20{,}000°$, $\beta_{12} = 0.001$ and two values of the dilution coefficients, $W = 0.1$ (case I) and $W = 0.01$ (case II).

TABLE 1

Line	Case I	Case II	Nebula
H_α	1.61	0.97	2.59
H_β	1.00	1.00	1.00
H_γ	0.44	0.80	0.50
H_δ	0.24	0.50	0.30
H_ε	0.15	0.32	0.19

The same table gives for comparison the Balmer decrement for transparent envelopes (e.g., nebulae) with $T_e = 20{,}000°$. They were obtained from Eqs. (15), which are a particular case of Eqs. (16) for $\beta_{1i} = 0$, $\beta_{ik} = 1$ ($i = 2, 3, \ldots$) when ionization from excited states is ignored.

The Balmer decrement for nebulae depends only on T_e and changes very slowly with its variation. In envelopes with a velocity gradient, the Balmer decrement is a function of several parameters and may take on widely scattered values.

Observations show that in the spectra of Wolf—Rayet, P Cygni, Be stars and novae the Balmer decrement changes markedly from one star to another, whereas in the spectra of Be stars and novae it is even variable within each individual spectrum This can be associated with possible variation of the parameters W and β_{12} in the envelopes of these stars, as the Balmer decrement is highly sensitive to their values.

Calculation of β_{12} requires knowledge of the velocity field in the envelope. Suppose that the atoms move in radial direction with velocity v, which is a function of r. It is readily seen that in this case

$$\frac{\partial v_s}{\partial s} = \frac{dv}{dr}\cos^2\vartheta + \frac{v}{r}\sin^2\vartheta, \tag{25}$$

where ϑ is the angle between the radius-vector and the line of sight.

Eq. (25) shows that even when $\frac{dv}{dr} = 0$, there is a velocity gradient in the envelope (owing to its curvature). In this case

$$\overline{\frac{\partial v_s}{\partial s}} = \frac{2}{3}\frac{v}{r}. \tag{26}$$

Having found $\overline{\frac{\partial v_s}{\partial s}}$, we can determine β_{12} from

$$\beta_{12} = \frac{1}{2un_1k_{12}}\left|\overline{\frac{\partial v_s}{\partial s}}\right|, \tag{27}$$

where n_1 is the number of atoms in the ground state in 1 cm^3, k_{12} is the absorption coefficient in the resonance line for one atom.

Estimation of β_{12} from the above equations gives the values required for the interpretation of the observed Balmer decrement.

§107. Be stars. The characteristic line shapes of the Be spectra (see Figure 149) are currently attributed to the fast spin of these star, which causes intermittent ejection of stellar matter. The spin of the star broadens the absorption lines which originate in the atmosphere. Judging from the line widths, the rotation speeds of Be stars reach hundreds of km/sec. The emission lines superimposed on the wide absorption bands originate in the envelope ejected from the star proper; they are produced by photoionization and recombination reactions.

The spin of the star is apparently conducive to ejection of stellar matter. The rotation, however, is hardly the main reason for this ejection, since the phenomenon is irregular.

The emission line profiles calculated from the equations of §105 for the case of gas ejection from a spinning star are similar to the line profiles observed in the spectra of Be stars. The variation of line profiles may be attributed to changes in the ejection power. This, in particular, affects the relationship between the transparent and the opaque parts of the envelope. Gorbatskii used this approach /7/ to interpret the variation of line profiles in the spectrum of γ Cas, one of the best known Be stars.

The emission line intensities in the spectra of Be stars were measured by various authors /8, 9, 10/. On the whole, the observed Balmer decrement does not fit the calculated decrement for nebulae. This can be remedied, however, using the theory of §106, which takes into consideration the opacity of envelopes with a velocity gradient. The inevitability of introducing envelope opacity for the interpretation of the Balmer decrement in the spectra of Be stars was also noted by observers /10/.

The continuous spectra of Be stars were studied by Barbier and Chalonge /11/. They measured the spectrophotometric temperatures of the stars up to the Balmer limit and above it; they also measured the Balmer jump. The results show that Be stars (B-type with emission) are on the average redder than B-type stars without emission. Moreover, the Balmer jump of Be stars proved to be less than the jump in B-type stars.

This difference in the continuous spectra of B and Be stars is associated with the presence of envelopes in the latter. The high-frequency radiation degraded in the envelope produces continuum quanta, as well as line quanta.

The theory of the continuous spectrum of Be stars is attractively simple /12, 13/. Since the stellar envelope is transparent at the continuum frequencies (the star is visible through the envelope), the energy emitted by a Be star at the frequency ν can be written in the form

$$L_\nu = L_\nu^* + L_\nu^{env}, \tag{28}$$

where L_ν^* is the energy emitted by the star proper (without the envelope), and L_ν^{env} is the energy emitted by the envelope.

Using Eq. (79) of the preceding chapter, which determines the quantity of energy emitted in 1 cm^3 in 1 sec, we get

$$L_\nu^{env} = \int n_e n^+ \frac{2^7\pi^3}{(6\pi)^{3/2}} \frac{e^6}{m^2c^3} \left(\frac{m}{kT_e}\right)^{\frac{1}{2}} \left(1 + 2\frac{\chi_1}{kT_e} \sum_{i=j}^{\infty} \frac{1}{i^3} e^{\frac{\chi_i}{kT_e}}\right) e^{-\frac{h\nu}{kT_e}} dV. \tag{29}$$

where the integral is taken over the entire envelope. The variation of n_e and n^+ in the envelope is determined by analyzing the transfer of ionizing radiation through the envelope. If the envelope is completely ionized (the optical thickness beyond the Lyman limit is less than unity) and the density of matter is inversely proportional to the square of the distance from the star center, Eq. (29) takes the form

$$L_\nu^{env} = 4\pi r_0^3 n_0^2 \frac{2^7\pi^3}{(6\pi)^{3/2}} \frac{e^6}{m^2c^2} \left(\frac{m}{kT_e}\right)^{\frac{1}{2}} \left(1 + 2\frac{\chi_1}{kT_e} \sum_{i=j}^{\infty} \frac{1}{i^3} e^{\frac{\chi_i}{kT_e}}\right) e^{-\frac{h\nu}{kT_e}}, \tag{30}$$

where r_0 is the inner radius of the envelope (equal to the radius of the star) and n_0 is the number of hydrogen atoms in 1 cm^3 on the inner surface.

It follows from the preceding equations that as the ejection of matter from the star is intensified, its light should show the following changes:

1) the apparent magnitude of the star should decrease;

2) the spectrophotometric temperature should drop (since the spectrophotometric temperature of the envelope is low);

3) the Balmer jump should become smaller (since the Balmer jump of the envelope is negative).

The last two conclusions are based on Eqs. (80) and (82) of the previous chapter.

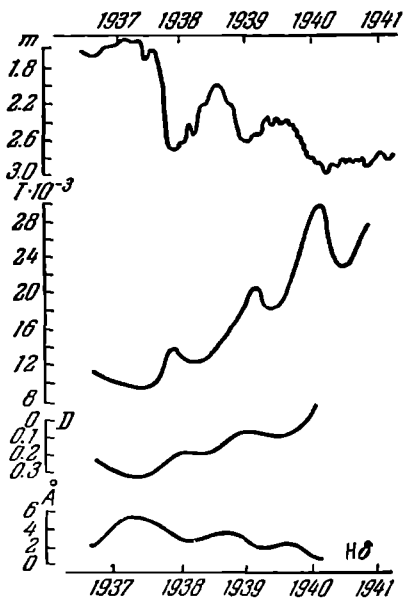


FIGURE 150

The above changes in the continuous spectrum are accompanied by an increase in the energy radiated by the envelope in the line spectrum. All these effects become more pronounced as the temperature of the star and the optical thickness of the envelope beyond the Lyman limit increase.

The above equations can also be used to obtain a detailed interpretation of the continuous spectrum of Be stars. For γ Cas this was done by Gorbatskii /7/. Between 1936 and 1941 the magnitude and the spectrum of this star underwent very considerable changes. The luminosity increased three times during this period. At the same time the spectrophotometric temperature decreased, the Balmer jump became smaller, and the Balmer line intensities increased (Figure 150). All these phenomena can be understood if we assume that three times during this period the ejection of stellar matter increased and subsequently subsided. In other words the star has shed three successive envelopes. By comparing the theory with observations, Gorbatskii determined the fundamental parameters of the star and the envelope. He found that the radius of the star was equal to three solar radii, the temperature of the star 34,000°, the number of hydrogen atoms in 1 cm^3 of stellar surface $n_0 \approx 10^{12} - 10^{13}$ and the average electron temperature of the envelope $T_e \approx 15{,}000 - 20{,}000°$.

§108. Wolf—Rayet stars. Since the publication of Beals's study /1/, it is generally held that Wolf—Rayet stars continulusly eject gaseous matter. This point of view is based largely on the profiles of spectral lines. The line profiles in the spectra of Wolf—Rayet stars are on the whole similar to the line profiles in the spectra of novae. This fact distinctly proves that Wolf—Rayet stars, like novae, eject stellar matter. The spectra of novae, however, change with time, which is due to the steady recession of the ejected envelope from the star. The spectra of Wolf—Rayet stars, on the other hand, are unchanging, which suggests that these stars eject gaseous matter continuously and at a constant rate.

The theoretical line profiles determined assuming continuous ejection of stellar matter are also similar to the line profiles in Wolf—Rayet spectra. The emission lines originate throughout the extended atmosphere whereas the violet absorption satellites of the emission bands form in the part of the atmosphere approaching the observer. Even if the atmosphere is opaque to the radiation in a given line, the energy emitted in this line by the receding part of the atmosphere will reach the observer because of the Doppler shift. The weakness of the absorption lines in Wolf—Rayet spectra can be attributed either to the small optical thickness of their atmospheres in these lines (because of the high ionization of the atoms) or to superposition of emission lines on the absorption lines. However, if the lower level is metastable, the absorption line is very strong. This is due to the accumulation of atoms in the metastable states, as discussed in §99. An example is provided by the absorption line λ 3889 Å which is associated with the 2^3S metastable state of neutral helium (see Figure 148).

A particularly important argument in favor of the ejection hypothesis for Wolf—Rayet stars is provided by the observed proportionality between the emission line width $\Delta\lambda$ and the wavelength λ. This is precisely the dependence predicted theoretically for lines originating in an expanding envelope. Table 2 lists an an example the measured and the calculated line widths $\Delta\lambda$ for three Wolf—Rayet stars (the proportionality coefficient between $\Delta\lambda$ and λ is given under the star designation). The gas ejection velocities obtained from emission line widths for Wolf—Rayet stars are of the order of 1,000—2,000 km/sec.

TABLE 2

λ	HD 192163, k=0.00833		HD 50896, k=0.00961		HD 191765, k=0.00883	
	$\Delta\lambda_{meas}$	$\Delta\lambda_{calc}$	$\Delta\lambda_{meas}$	$\Delta\lambda_{calc}$	$\Delta\lambda_{meas}$	$\Delta\lambda_{calc}$
6563	58.0	54.7	74.2	63.1	57.5	57.9
4861	41.2	40.5	44.8	46.7	43.9	42.9
4340	33.4	36.2	36.0	41.7	37.8	38.3

Recently discovered binaries with Wolf—Rayet components also provide evidence in favor of the hypothesis of gas ejection. The results of photometric and spectroscopic observations of these binary systems indicate that the companions of Wolf—Rayet stars move in extended atmospheres in which the principal stars are shrouded /14, 15/.

Finally note that the hypothesis of gas ejection from Wolf—Rayet stars associates the emission bands in their spectra with the fluorescence mechanism. As we know, fluorescence requires small dilution coefficients. This condition is indeed satisfied in extended atmospheres. If we were to reject the notion of extended atmospheres around Wolf—Rayet stars, the interpretation of the bright bands in their spectra would present considerable difficulty.

It is significant that the assignment of the emission bands in Wolf—Rayet spectra to photoionization and recombination processes is also supported by a number of independent facts. The extended atmospheres of Wolf—Rayet stars are in fact assumed to show the same stratification as the

planetary nebulae. This means that lines of high ionization potentials originate in deeper lying layers of the atmosphere than lines of low ionization potential do. In planetary nebulae the stratification of radiation can be directly observed, whereas for Wolf—Rayet stars it is unobservable. The stratification of radiation in the atmospheres of Wolf—Rayet stars nevertheless emerges from the following considerations.

As we know, the nuclei of some planetary nebulae are Wolf—Rayet stars. In these cases, Zanstra and Weenen /16/ note that if the spectrum of the nucleus contains He II and He I lines, the spectrum of the nebula shows only He I lines; if, however, the spectrum of the nucleus contains only He II lines, and no He I lines, the spectrum of the nebula shows both He I and He II lines. This effect is associated with the stratification of radiation in the atmospheres of Wolf—Rayet stars. In the former case the entire stellar energy beyond the limit of the He II principal series is absorbed in the stellar atmosphere, whereas in the latter case the stellar atmosphere absorbs only part of this energy. In other words, in the first case the boundary between the He II and He I emitting regions lies in the stellar atmosphere, and in the second case it lies in the nebula.

TABLE 3

Stars	He I	He II
HD 192 163	1815	1550
151 932	735	530
177 230	785	520
192 103	1290	975
184 738	1075	540

Another indication of the stratification or radiation in the atmospheres of Wolf—Rayet stars is provided by the unequal widths of the emission lines correspond to different atoms in these spectra. For example, the He I lines are wider than the He II lines. This is evident from Table 3, which lists for various Wolf—Rayet stars the line widths of He I (λ 5875 Å) and He II (λ 5411 Å), expressed in km/sec.

The different widths of He I and He II lines are associated with their formation in different parts of the atmosphere, where the velocities of atoms are a function of the distance from the star's center. Since the He I lines originate in upper lying layers (compared to He II lines), Table 3 shows that the motion of atoms in the atmospheres of Wolf—Rayet stars is accelerated. Minin's calculations show that the accelerated motion of helium atoms ejected from a Wolf—Rayet star can be caused by radiation pressure beyond the limits of the principal He II and He I series.

The recombination mechanism of emission in the line spectra of Wolf—Rayet stars led some authors to certain conclusions concerning the chemical composition and the physical conditions in stellar atmospheres, based on line intensities.

Ambartsumyan /17/ determined the relative content of hydrogen and helium in the atmosphers of Wolf—Rayet stars from the intensity ratio of H_β and λ 4686 Å He II. He treated the atmospheres as transparent to line radiation. The atmospheres of Wolf—Rayet stars, unlike other atmospheres, contain more helium than hydrogen (approximately by a factor of 1.8).

Zanstra and Weenen /16/ also assumed transparency of the atmospheres in the relevant lines and determined the relative abundance of hydrogen, helium and carbon. First they calculated the size of the regions which emitted the lines corresponding to different ionization stages of the atoms. The helium to hydrogen ratio was found to be 1.3, and the carbon to helium ratio 0.37.

Recently Slyusarev /18/ considered the intensity of HI and He II emission lines in the spectra of Wolf—Rayet stars using the theory of §106. The atmospheres of these stars were found to be transparent to the radiation in hydrogen lines and opaque to He II radiation. The transparency of the atmospheres in hydrogen lines is associated with the high ionization of hydrogen and the large velocity gradients in the atmospheres. The He II line intensities calculated from Eqs. (16) showed good fit with observations. The HI/He II intensity ratio indicated that the number of helium atoms in the atmosphers of Wolf—Rayet stars was a factor of 5—12 greater than the number of hydrogen atoms.

It follows from the above studies that the chemical composition of Wolf—Rayet stars is highly anomalous. In the atmospheres of "normal" stars and in gaseous nebulae the abundance of hydrogen is about 10 times as great as the abundance of helium, whereas in Wolf—Rayet stars the ratio is reversed. This point deserves special attention, since according to modern concepts the energy of the stars is generated in nuclear reactions which convert hydrogen into helium.

The intensity ratio of the emission lines to the continuous spectrum can be used in determining the temperatures of Wolf—Rayet stars. This is done applying Eq. (24) of the previous chapter to different atoms. Table 4 lists the temperatures of Wolf—Rayet stars (in thousands of degrees) calculated by Vorontsov-Vel'yaminov /19/. In each column, the ionization potentials in volts are indicated below the atomic symbol.

TABLE 4

Stars	He I 24.5	C III or N III 47.4	He II 54.2	C IV 64.2	N IV 77.0	T_{sp}
HD 192 163	32	65	73	—	84	15
191 765	35	62	69	—	75	15
193 077	29	51	59	—	74	13
193 576	29	48	60	—	62	14
192 103	33	64	63	69	—	12
192 641	—	59	55	70	—	7

We see that Zanstra's method gives extremely high values for the temperatures of Wolf—Rayet stars. The table also shows that the lines of different atoms give different temperatures for the same star. This is mainly associated with the small optical thickness of the atmospheres beyond the limit of the principal series of atoms with low ionization potentials, so that the atmosphere absorbs only part of the stellar radiation at these frequencies. This is the case with neutral helium and hydrogen. Calculations show that the optical thickness of the atmosphere of a Wolf—Rayet star beyond the Lyman limit is of the order of 0.01. The temperatures determined from hydrogen lines are therefore very low, about 20,000°. Another reason for the divergence between the temperatures obtained from different lines is the deviation of the actual energy distribution in the stellar spectrum from the Planckian distribution.

The last column in Table 4 lists the spectrophotometric temperatures of Wolf—Rayet stars. They are clearly much lower than the temperatures

obtained by Zanstra's method. This is so because the degradation of high-frequency radiation in the envelopes of Wolf—Rayet stars produces not only emission lines but also a continuous spectrum with energy distribution corresponding to an exceedingly low temperature. The processes producing the continuous spectrum of Wolf—Rayet stars, however, are more complex than the corresponding processes for Be stars. This is associated with the greater thickness of the Wolf—Rayet envelopes, which therefore act both as an "atmosphere" and a "photosphere" (the star "proper" does not show throught the envelope).

Note that the energy emitted by a Wolf—Rayet star in a line spectrum can be used to determine the concentration of atoms at the atmosphere—photosphere interface. It is found to be of the order of $10^{12}\,\mathrm{cm}^{-3}$.

Of great interest is the determination of the quantity of matter ejected by a Wolf—Rayet star annually. Clearly, the discharge of matter through $1\,\mathrm{cm}^2$ area at a distance r from the star's center in time dt is $\varrho(r)v(r)dt$, where $\varrho(r)$ and $v(r)$ are respectively the density and the velocity at a given point in the envelope. The discharge through the entire spherical surface of radius r in time dt is

$$dM = 4\pi r^2 \varrho(r)\, v(r)\, dt. \tag{31}$$

To evaluate $\frac{dM}{dt}$, we take r to be the radius of the photosphere. Taking $r \approx 5r_\odot$, $\varrho \approx 10^{-11}\,\mathrm{g/cm}^3$ (this ϱ corresponds to the previously mentioned concentration of helium atoms, $10^{12}\,\mathrm{cm}^{-3}$), $v = 10^8\,\mathrm{cm/sec}$, we find that a Wolf—Rayet star loses annually about 10^{-5} of the solar mass.

This result is of great significance in cosmogony. Since the masses of Wolf—Rayet stars are of the order of 10 solar masses, a star cannot remain for more than 10^6 years in the Wolf—Rayet stage. Now, as there are apparently no stars with larger masses (except the related O stars), we conclude that the Wolf—Rayet stars formed directly from the proto-stellar matter in the very recent past.

Note that the extreme youth of Wolf—Rayet stars is also borne out by other data. Ambartsumyan has recently discovered a new type of star systems /20/, stellar associations. These systems are characterized by high density of stars of certain spectral types (O and B stars in O associations, and T Tauri stars in T associations). According to Ambartsumyan's calculations, stellar associations are unstable, rapidly decaying formations. Hence the conclusion that these associations formed not long ago. The age of the associations is a few million years. Since most of the Wolf—Rayet stars are found in associations, they are also young objects.

Eq. (31) can be used to estimate the annual mass ejection of P Cygni and Be stars. It is found to be of the order of $10^{-5} M_\odot$ for P Cygni stars and of the order of 10^{-6}—$10^{-8} M_\odot$ for Be stars. These stars, like Wolf—Rayet stars, are typical members of stellar associations.

§109. Late-type stars with bright lines. Wolf—Rayet, P Cygni and Be stars are not the only ones with bright lines in their spectra: some late-type stars, such as long-period variables, Z Andromedae stars, and others also show emission lines.

The magnitude and the spectrum of the long-period variables vary with periods of the order of 1 year. The range of light variation is several

stellar magnitudes. Near the maximum, the spectrum shows bright lines of hydrogen and ionized iron, and near the minimum bright lines of neutral iron. Most of the long-period variables are of spectral type M, and of these nearly 80% contain bright emission lines in their spectra.

The spectra of Z And stars are a combination of a late-type spectrum with absorption lines and an early-type spectrum with emission lines corresponding to atoms of high ionization potentials (e.g., He II). The magnitude and the spectrum of these stars show irregular variations. Besides Z And, this group also includes R Aqr, V Sge, and others.

The main question to be considered at this point is the actual reason for the appearance of emission lines in the spectra of these late-type cold stars. As we know, the emission lines in the spectra of hot stars are the result of the degradation of the high-frequency stellar radiation in an extended envelope. However, the high-frequency energy of late-type stars is too weak to ensure adequate emission lines through this mechanism.

The spectra of Z And stars are generally interpreted assuming that these are binaries with one hot and one cold component. The light variation curves of the hypothetical components were actually determined. Some of these stars are possibly binaries, but we have no conclusive proof of this. The hypothesis is moreover inapplicable to long-period variables.

An alternative hypothesis regards the late-type stars with emission lines as hot stars with extended envelopes of great optical thickness in the continuous spectrum /12/. We have seen above that for Be and Wolf—Rayet stars the temperatures deduced from the bright lines are substantially greater than the spectrophotometric temperatures. As the optical thickness of the envelope increases, this disparity in temperatures should increase. If we assume an envelope of extremely large optical thickness (this is so if the density in the envelope decreases slowly with increasing distance from the star's center), the following physical processes become possible. The interior of the envelope will absorb almost all the energy emitted by the star, degrading it into low-frequency quanta. This produces a late-type continuous spectrum and emission lines whose intensity corresponds to the temperature of the star proper. In the outer parts of the envelope, which are mainly exposed to low-temperature radiation from the interior, neutral metals and molecular compounds are found. This skin gives the late-type absorption spectrum.

The above hypothesis is on the whole borne out by observations. It is particularly significant that the emission lines observed in the spectra of long-period variables originate in deep-lying parts of the envelope, whereas the absorption lines form in the outer layers. This follows from the fact that hydrogen line emission is partly absorbed in the stellar atmosphere. Some of the Balmer lines are split into several components, which is associated with the absorption of radiation in these lines by the atoms of neutral metals. The spectra of Me stars reveal an unusual distribution of intensities among the Balmer lines. Shain /21/ has shown that this intensity distribution is associated with the absorption of hydrogen radiation in titanium oxide bands. The spectra of Ne and Se stars show no titanium oxide bands, and the Balmer decrement is quite normal there.

The formation of emission and absorption lines at different depths in the envelope also follows from the different behavior of the radial velocity curves obtained from these lines. Note that the K-term found from emission

lines is positive, being approximately equal to 15 km/sec, whereas the K-term determined from absorption lines is close to zero. This leads to the conclusion that gaseous matter is ejected from the long-period variables and that it is greatly slowed down at large distances from the star. The deceleration of ejected matter may account for the slow decrease of density in the shell.

If the optical density of the envelope becomes extremely high, no emission lines will be observed. The "normal" cold supergiants probably form in this way. The masses and the luminosities of M-type supergiants are nearly those of O and B stars. This fact alone suggests that the stars of these different types differ only in the structure of their envelopes.

According to this hypothesis, Z And stars are O-type stars with envelopes in which the density decreases slowly along the radius. The optical thickness of these envelopes in the continuous spectrum is nevertheless exceedingly large. We therefore observe "combined" spectra, which are in fact a superposition of two spectra, that of the star and that of the envelope. Variation in magnitudes and spectra of these stars may be associated with changes in the ejection rate of stellar matter.

In addition to the various high-luminosity stars, there are also some late-type dwarfs with emission lines in their spectra. The most remarkable in this group are the T Tauri variables. These are stars of spectral types G—M with H, Ca II, Fe II, and other emission lines. The bright lines are accompanied by violet absorption satellites. Judging from the line profiles, we conclude that T Tau stars eject gaseous matter.

Almost all the known T Tauri stars occur in stellar associations (so-called T associations). According to Ambartsumyan, this is indicatory of their extremely low age.

Recently Ambartsumyan /22/ called attention to a remarkable feature of T Tauri and related stars. In the epoch of light enhancement, the appearance and enhancement of emission lines entails the development of a very strong continuous spectrum, which is superimposed on the normal continuous spectrum with absorption lines. Since the new continuous spectrum attenuates all the absorption lines without producing any new ones, it seems to originate in the topmost layers of the stellar atmosphere. This conclusion is also borne out by the simultaneous appearance of emission lines with the continuous spectrum.

In UV Ceti stars the flare is only a few minutes long, and the luminosity increases by a few stellar magnitudes. This rapid and conspicuous increase in the star's luminosity cannot be associated with influx of energy transported from the stellar interior to the surface by heat conduction or radiative transfer. We therefore conclude that the energy radiated by the UV Ceti flare stars is delivered to the outer atmospheric layers by some hitherto unknown mechanism.

Examination of various stars with additional emission in the continuous spectrum led Ambartsumyan to the conclusion that we are invariably dealing with ejection of some stellar matter — the source of stellar energy — from the interior. The stellar matter ejected from the interior apparently undergoes rapid decay, not unlike nuclear disintegration, and releases substantial quantities of energy which is eventually converted to visible radiation. It would seem that disintegration of stellar matter also produces

atoms of various elements. In this way Ambartsumyan explains the presence of technetium, deuterium, and other short-lived atoms in stellar atmospheres.

Note that short-lived atoms also occur in the solar atmosphere. Ambartsumyan maintains that interior matter is ejected even by Sun-type stars, although this phenomenon takes place on a smaller scale than in T Tauri and other related star. The large scale of these phenomena in T Tauri stars is attributed by Ambartsumyan to the early age of these stars. In his opinion, the ejected matter is the protostellar gas conserved in the stellar interior.

REFERENCES TO CHAPTER XVII

1. BEALS, C.S. Spectrophotometric studies of Wolf — Rayet stars and novae.— Publ. Obs. Victoria 6, No.9, 1934.
2. CHANDRASEKHAR, S. On the hypothesis of the radial ejection of high speed atoms from the Wolf— Rayet stars and the novae stars.— MN 94:522. 1934.
3. GERASIMOVIC, B. The contours of emission lines in expanding nebular envelopes.— Zs. f. Astrophys., 7:335. 1933.
4. SOBOLEV, V.V. Shapes of spectral lines forming in moving stellar atmospheres.— Astron. Zh., 25:3. 1948,
5. AMBARTSUMYAN, V.A., E.R. MUSTEL, A.B. SEVERNYI, and V.V. SOBOLEV. Teoreticheskaya astrofizika (Theoretical Astrophysics), Ch. VI.— Gostekhizdat. 1952.
6. SOBOLEV, V.V. Intensity of emission lines in the spectra of early-type stars.— Astron. Zh., 24:13. 1947.
7. GORBATSKII, V.G. Interpretation of the spectral variations of γCassiopei.—Astron. Zh., 26:307. 1949; 28:450. 1951.
8. KARPOV, B.G. A study of hydrogen emission in stars of class B0—B5.— Lick. Obs.Bull., 16:159. 1934.
9. MOHLER, O. A determination of the temperatures of Be stars.— Publ. Obs. Univ. Michigan, 5:43. 1934.
10. BURBIDGE, G.R. and E.M. BURBIDGE. The Balmer decrement in some Be stars.— ApJ 118:252. 1953.
11. BARBIER, D. and D. CHALONGE. Intensites totales des raies de l'hydrogene et du calcium ionise dans les spectres stellaires.— Ann. d'Astrophys., 4:1. 1941.
12. SOBOLEV, V.V. Dvizhushchie obolochki zvezd (Moving Stellar Envelopes), Ch. V.— Leningrad University. 1947.
13. GORBATSKII, V.G. Ionization in extended stellar envelopes.— Uchenye Zapiski, Leningrad University, No.136. 1950.
14. WILSON, O. Absolute dimensions of a Wolf— Rayet star and the expanding envelope hypothesis.— ApJ 95:402. 1942.
15. BEALS, C. Some results of a spectrophotometric study of the Wolf —Rayet binary HD 193576.— MN 104:205. 1944.
16. ZANSTRA, H. and J. WEENEN. Wolf—Rayet stars and Beals' hypothesis of pure recombination.— Bull. Astr. Inst. Netherlands, XI, No.411, 1950.
17. AMBARTSUMYAN, V.A. On the intensities of λ 4686 and H_β in the Wolf — Rayet stars.— Tsirkulyar, Pulkovo Observatory, No.7. 1933.
18. SLYUSAREV, S.G. Relative content of hydrogen and helium in the atmospheres of Wolf—Rayet stars.— Astron. Zh., 32:346. 1955.
19. VORONTSOV-VEL'YAMINOV, B.A. Intensity and photometric effect of bright bands in the spectra of six Wolf—Rayet stars and their temperature.— Astron. Zh., 23:1. 1946.
20. AMBARTSUMYAN, V.A. Stellar associations.— Astron. Zh., 26:3. 1949.
21. SHAIN, G.A. Uber die Emissionslinien des Wasserstoffs in den Spectren langperiodischer Veranderlicher.— Zs. f. Astrophys., 10:73. 1936.
22. AMBARTSUMYAN, V.A. The phenomenon of continuous emission and sources of stellar energy.— Bull. Byurakan Obs., No. XIII. 1954.

Part Six

STELLAR SYSTEMS

V. V. SOBOLEV

Chapter XVIII

STELLAR STATISTICS. GALACTIC STRUCTURE

§110. **Galactic concentration.** Stellar statistics emerged as a science in the far antiquity, when Shih Sheng (fourth century B. C.) and Hipparchus (second century B. C.) compiled the first star catalogues.

The initiator of modern stellar statistics was William Herschel, at the end of the 18th century. By counting the stars in the field of view of his telescope in different parts of the sky, Herschel discovered one fundamental fact, namely that the number of stars per unit sky area increases as one approaches the Milky Way. This phenomenon is known as galactic concentration. It was subsequently established that galactic concentration is a characteristic feature of all the objects in our Galaxy. The Milky Way itself is a manifestation of galactic concentration.

Herschel correctly interpreted the galactic concentration as a result of oblateness of our stellar system. The plane through the celestial circle toward which the observed objects concentrate was identified by Herschel with the plane of symmetry of the stellar system. This plane is known as the galactic plane, and the corresponding large circle is the galactic equator. Stellar motions also display certain symmetry relative to the galactic plane; this is understandable, as stellar motions and positions are interrelated.

Since the galactic plane is a plane of symmetry for both stellar positions and stellar motions, the coordinate system traditionally used in stellar astronomy adopts this plane as the reference plane, and the galactic circle is thus the reference circle.

§111. **Galactic coordinates.** To define a galactic system of coordinates, we require the position of the galactic equator in the equatorial coordinate system. We must therefore formulate in precise terms the concept of the galactic equator. Its definition as the circle in the galactic plane of symmetry is not rigorous enough, since the distribution of various objects over the sky is not strictly symmetric.

Mathematically, the problem of determining the position of the galactic equator can be formulated as follows: consider any great circle on the celestial sphere, defined by its inclination i to the celestial equator and the right ascension of the ascending node Ω. Determine the angular distance $x_k(i_m, \Omega_m)$ of all galactic objects up to some given magnitude m from this great circle. Calculate the sum $G_m = \sum_{k=1}^{n_m} x_k^2(i_m, \Omega_m)$, where n_m is the number

of all galactic objects up to magnitude m. Find the values of i_m and Ω_m which minimize the sum G_m. The galactic equator is then defined as the great circle whose inclination and the right ascension of the ascending node are the limit values of i_m and Ω_m as $m \to \infty$.

In practice the position of the galactic circle is often determined graphically: it is drawn through the region of maximum concentration of certain selected populations of objects. This approach clearly gives similar, but not identical results. Yet the galactic system of coordinates must be defined in a precise and unique fashion. By convention the following standard values of the inclination and the right ascension of the ascending node have been adopted: $i = 62°$, $\Omega = 280°$. These values are the basis for Ohlsson's conversion tables from equatorial to galactic coordinates (Annals of the Observatory of Lund, No.3. 1932), which are in common use in stellar astronomy. The corresponding equatorial coordinates of the galactic pole are $A = 190°$ and $D = 28°$ (epoch 1900.0). These coordinates were adopted as the standard by the International Astronomical Union in 1935. Later determinations, based mainly on radio observations, gave somewhat different results. The General Assembly of the International Astronomical Union in 1958 therefore adopted a new standard system of galactic coordinates, which is defined by the equatorial coordinates of its pole $A = 12^h49^m$, $D = +27°.4$ (epoch 1950.0). In 1961 the Observatory of Lund published another set of conversion tables for this position of the galactic pole (Annals, Nos. 15, 16, 17). The galactic system does not change its orientation in space and is independent of the observation epoch.

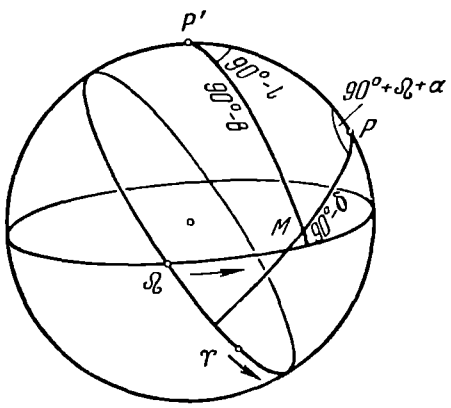


FIGURE 151

The galactic latitude is designated b. The galactic longitude l is reckoned counterclockwise from the ascending node (the point of intersection of the celestial equator with the galactic equator where the galactic equator crosses from the southern to the northern celestial hemisphere, if one goes in the direction of increasing galactic longitudes) (Figure 151).

The galactic and the equatorial coordinates of a star are related by the standard relations of spherical trigonometry:

$$\left.\begin{aligned} \cos b \cos l &= \cos(\alpha - \Omega)\cos\delta, \\ \cos b \sin l &= \sin i \sin\delta + \cos i \sin(\alpha - \Omega)\cos\delta, \\ \sin b &= \cos i \sin\delta - \sin i \sin(\alpha - \Omega)\cos\delta, \end{aligned}\right\} \qquad (1)$$

$$\left.\begin{aligned} \cos\delta\cos(\alpha - \Omega) &= \cos b \cos l, \\ \cos\delta\sin(\alpha - \Omega) &= -\sin b \sin i + \cos b \cos i \sin l, \\ \sin\delta &= \sin b \cos i + \cos b \sin i \sin l. \end{aligned}\right\} \qquad (2)$$

These relations are easily derived by considering the spherical triangle with its vertices at the celestial pole P, the galactic pole P', and the star M. To convert from equatorial to galactic coordinates, the former should first be referred to the epoch 1900.0.

The ascending node Ω is located in the constellation of Aquilla, the galactic north pole in Coma Berenices, and the south pole in Sculptor.

§112. Luminosities and absolute stellar magnitudes. Spectrum — luminosity diagram. The luminosity L of a star is the quantity of radiant energy emitted by the star in unit time. L is generally expressed in units of solar luminosity. The radiation flux H at a given point is the quantity of radiant energy crossing in unit time a unit surface element at right angles to the direction of propagation of the radiation. Since stars are spherical objects, radiating uniformly in all directions, the luminosity of a star and the radiation flux at a given point (if the interstellar space is perfectly transparent) are related by the equation

$$H = \frac{L}{4\pi r^2}, \tag{3}$$

where r is the distance, in parsecs, from the star. Here atmospheric absorption is also ignored.

As we know, the apparent stellar magnitude m is expressed in terms of the luminous flux by the relation

$$m = -2.5 \lg H + c, \tag{4}$$

where c is a constant which determines the zero point of the scale of stellar magnitudes. The absolute stellar magnitude M is defined as the apparent magnitude that the star would have at a distance of 10 pc, given a perfectly transparent interstellar space and absorption-free atmosphere. Then

$$M = -2.5 \lg \frac{L}{4\pi \cdot 10^2} + c. \tag{5}$$

Subtracting (4) from (5) and using (3), we get

$$M = m + 5 - 5 \lg r \tag{6}$$

or, if p is the stellar parallax in seconds of arc,

$$M = m + 5 + 5 \lg p.$$

The absolute magnitude of a star is clearly independent of the observer's position and is thus a physical characteristic of the star.

The difference $m-M$, which in the absence of light absorption is equal to $5 \lg r - 5$, is called the distance modulus, as it provides a single-valued indication of the distance. To convert the distance modulus to the actual distance, it should be corrected for light absorption.

Another no less important physical charactieristic of stars is the spectral type or class. It was established by Hertzsprung in 1911 and independently by Russell in 1913 that the absolute magnitude of a star and its spectral type show a definite correlation. If the galactic stars are plotted in a diagram, laying the absolute magnitudes off the vertical axis and the spectral types off the horizontal axis, the result will be very much like the plot of

Figure 152. The position of the points in Figure 152 is according tc Parenago (1954), whose figures are the most reliable.

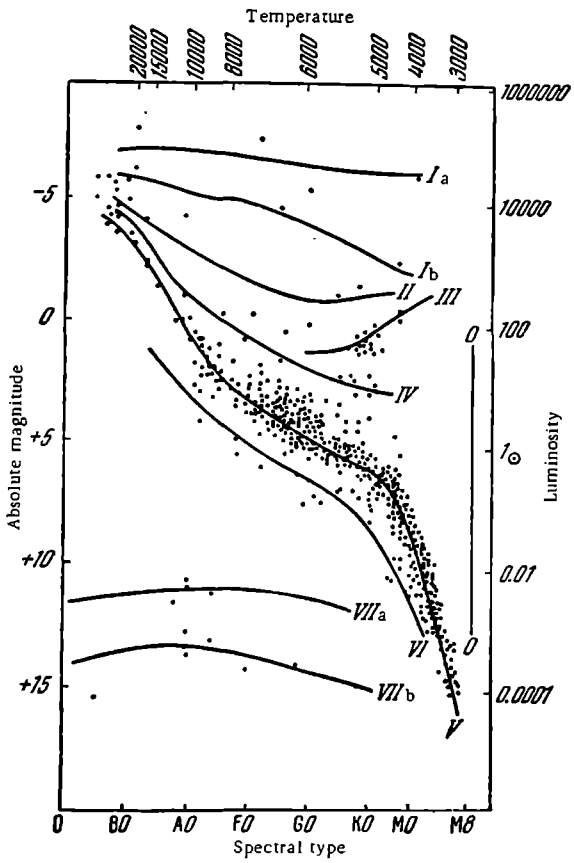


FIGURE 152

The points on this diagram cluster along several distinct lines. The currently used system of these lines (strips) was proposed by Morgan and co-workers, and they are also marked on Parenago's diagram. These nonintersecting lines, some of which are almost parallel to the abscissa axis whereas others are markedly inclined, reflect the fundamental fact that although stars of one spectral type may have different luminosities, the luminosity does not assume just any value within the permissible range but clusters around several fixed values. Therefore, following Morgan, the entire group of stars around each of the discrete lines in the diagram is classified as a separate luminosity class; this fact is denoted by Roman numerals (in some cases further qualified by the Latin letters a and b), marked near the corresponding lines in Figure 152.

The diagram shows that the bulk of stars fall in the diagonal strip (around the line marked V). This group of stars is called the m a i n s e q u e n c e. Above the main sequence lie three different lines marked

II, III, and IV. Stars falling in these parts of the diagram are respectively known as bright giants, normal giants, and subgiants. Previously all these giants were usually lumped into one group, and the corresponding non-differentiated part of the diagram was called the giant branch. Above the giant branch lie the less luminous supergiants (Ib) and still higher are the most luminous supergiants (Ia).

Some $2-2^m.5$ below the main sequence and almost parallel to it lies the subdwarf sequence (VI). Further down lie the bright white dwarfs (VIIa) and the weak white dwarfs (VIIb). The significance of the different sequences is growing progressively clearer from year to year, and therefore Morgan's notation assigning the Roman numeral of the sequence after the spectral type symbol, originally introduced in 1943, has now become quite popular. The different sequences are generally called luminosity classes. Thus, Betelgeuse is described as M2 I, Arcturus as K0 III, Sun as G2 V.

Stars situated in the bottom right-hand part of the main sequence are generally called red dwarfs. In many of the currently used catalogues (mostly compiled before the introduction of luminosity classes), each star, in addition to its spectral type, is classified as a normal (red) dwarf (d), a white dwarf (wd), a subdwarf (sd), a subgiant (sg), a giant (g), or a supergiant (c). The lower-case Latin letters given in parentheses precede the spectral type symbol. Thus, Betelgeuse is described as cM2, and Arcturus as gK0. Stars in the left-hand part of the main sequence and in the middle are neither dwarfs nor giants; they are described only in terms of their spectral type, e.g., G2 for the Sun.

In 1947 Vorontsov-Vel'yaminov established the existence of still another sequence, comprising high-temperature nonstationary stars. This so-called white-blue sequence extends from the white supergiants to the white dwarfs and includes Wolf—Rayet stars, nuclei of planetary nebulae, nova-like variables, and novae at their minimum.

Figure 152 shows that the most densely populated part of the spectrum — luninosity diagram, often called the Hertzsprung—Russel diagram, is the main sequence, especially in the red dwarf section. Giants and particularly supergiants are much rarer stars. The diagram shows giants from a very large volume of space, since these high-luminosity stars are visible over enormous distances, whereas the much fainter dwarfs are counted only over substantially smaller distances. This is the reason why the diagram seems to suggest that the giants are more numerous than the white dwarfs. In fact, the number of white dwarfs exceeds the number of giants by more than a factor of 100.

It seems that for every 10 million main sequence stars (mainly red dwarfs) there are some 1,000,000 white dwarfs, 10,000 subdwarfs, 1,000 giants, and only one supergiant. The number of subgiants is difficult to estimate even crudely. They are hardly more numerous than the giants.

The spectral type and the absolute magnitude of the stars figuring in the spectrum—luminosity diagrams are greatly distorted by measurement errors. The spectra of some stars are determined with very low reliability, especially if low-dispersion spectrograms are used. The errors in absolute magnitude are even larger, since they are associated with errors in distance determination and absorption corrections. We thus naturally

suspect that the spread of main-sequence stars around the central line may be an outcome of measurement errors. Conversely, the errors may prove to make but a negligible contribution and the scatter of the absolute stellar magnitudes of main sequence stars within each spectral type is significant.

To examine this question, Parenago plotted the spectrum—luminosity diagram of the nearest stars, whose trigonometric parallaxes are known with fair reliability and for which the absorption effect is negligible. In this plot, the vertical width of the main sequence strip shrank from $2-2^m.5$ to $\sim 1^m$. The residual scatter cannot be attributed to observation errors and is apparently significant.

§113. Color—luminosity diagram. The determination of the color indices CI of stars is a simpler undertaking than the determination of their spectral types. Taking spectra with a slit spectrograph is a lengthy procedure which must be repeated for every individual star. The color indices, on the other hand, can be determined "wholesale," by photographing a certain area of the sky once in photographic light and then again in photovisual light and taking the difference

$$CI = m_p - m_v \tag{7}$$

for each of the stars on the two plates. Moreover, the color index can be determined for much fainter stars (by several stellar magnitudes).

Therefore, for open and globular star clusters one generally plots the color index vs. absolute magnitude diagram; however, since the distance modulus is mostly unknown, a commonly used diagram plots the color index against the apparent magnitude.

The color index $m_p - m_v$ is an important characteristic of a star. However, the color index as such does not provide much information on the properties of stellar radiation. Using this parameter alone, we cannot distinguish, say, between a dwarf and a giant, or between a late-type star without interstellar absorption and an early-type star reddened by interstellar absorption.

To ensure more rational utilization of photometry, various systems of stellar magnitudes have been devised for different colors, i.e., stellar magnitudes in different spectral regions, defined by suitable filters. A discussion of these photometric systems will be found in Chapter I.

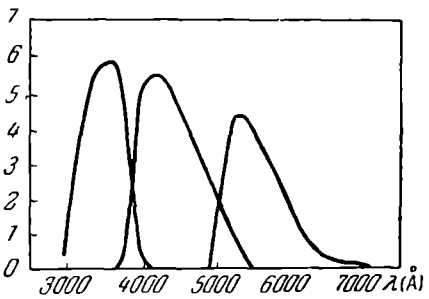


FIGURE 153

Stebbins and Whitford developed and applied six-color photometry. In 1954 Johnson and Morgan advanced their three-color photometry, which has now become highly popular. In this system, the stellar magnitudes are measured in three spectral regions, isolated by appropriate filters. Figure 153 shows the response curves of a photoelectric detector in Johnson and Morgan's three-color photometry. The apparent magnitude through a yellow filter, corrected for atmospheric absorption (to a value apparently

observed outside the atmosphere), is designated V. This magnitude is virtually equivalent to the photovisual magnitude in the International System. The apparent magnitude through a blue filter, again corrected for atmospheric absorption, is designated B. It incorporates a zero-point correction, so that a star of spectral type A0 V satisfies the equality $B-V=0$. The apparent magnitude through an ultraviolet filter, corrected for atmospheric absorption, is designated U. Here the zero point is also adjusted so that for stars of spectral type A0 V, $U-B=0$.

In the Johnson—Morgan system two parameters characterize the energy distribution in the star's spectrum. The first, $B-V$, is fairly close to the color index (7). The second, $U-B$, is also a sort of a color index, which characterizes the energy distribution in the short-wave region.

Johnson and Morgan picked out stars of luminosity classes III, V, and VII with reliable parallaxes (for trigonometric parallaxes $>0''.100$) and determined their absolute magnitudes in the color V (designated M_V). Furthermore, they postulated that stars of luminosity class V in the near open clusters of Pleiades and Praesepe have the same absolute magnitude as stars of luminosity class V and the same spectral type near the Sun. This enabled them to determine the distance modulus and hence the absolute magnitudes of the stars in these clusters. The result was a diagram (Figure 154) plotting the absolute magnitude against the color index for a number of red giants, main-sequence stars, and white dwarfs. This diagram was adopted as the standard for future tagging and construction of analogous diagrams for other groups of stars. Since all the stars of a given cluster have the same distance modulus, the corresponding diagram shows stars of all the sequences, and not only of luminosity classes III, V, and VII.

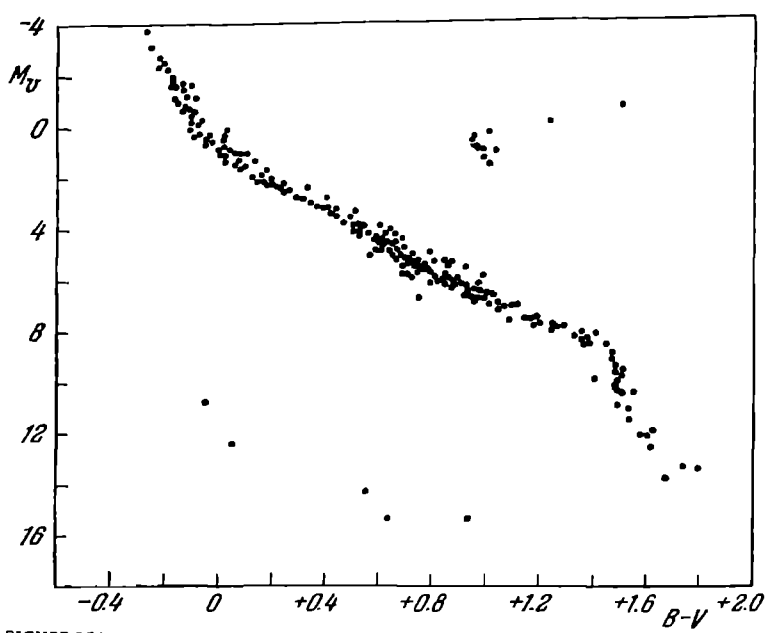


FIGURE 154

Comparison of this diagram with the spectrum—luminosity diagram (see Figure 152) reveals definite likeness in the position of the sequences III, V, and VII. This is a natural result, since the color index $B-V$ is closely related to the spectral type. The likeness, however, is not complete. For example, the giants are redder than the dwarfs of the same spectral type (their $B-V$ is higher), and therefore in the color—luminosity diagram the giant branch is somewhat displaced to the right compared to its position in the spectrum—luminosity diagram.

Between sequences V and III there is a substantial unfilled gap in the horizontal direction. This part of the diagrams, called the Hertzsprung—Russell gap, remains empty as long as we are considering stars in the spiral arms of the Galaxy (the neighborhood of the Sun included) or in open clusters, i.e., Population I stars (see §132).

Another remarkable feature is the small spread of the luminosities of the main sequence stars in Figure 154. This is probably associated with the high accuracy of the distances to the stars in this diagram (the scatter due to errors in distance determination is small). The small spread in luminosities possibly reflects great similarity of chemical composition and common origin of the stars. This is again associated with the mutual proximity of the stars in space.

Another standard diagram compares the $B-V$ and $U-B$ values of the nearest main-sequence stars with those of the main-sequence stars in the near open clusters Praesepe and NGC 2362 (Figure 155). This diagram shows that an increase in $B-V$ on the whole entails an increase in $U-B$.

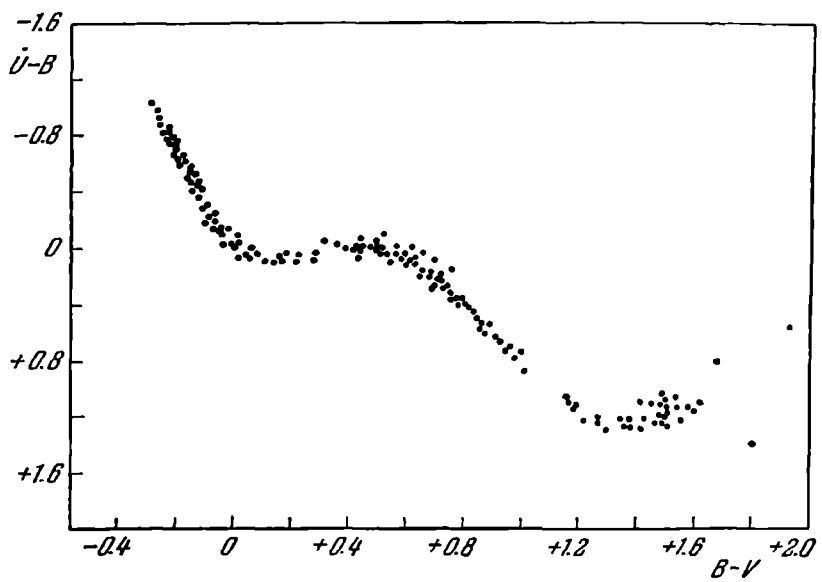


FIGURE 155

The strip is very narrow, except at its right-hand tail, where M-type dwarfs are located. The substantial spread of points in this region is

apparently associated with the presence of strong hydrogen emission lines in some M dwarfs (producing high-intensity ultraviolet radiation), while in other M dwarfs these lines are weak or absent altogether. The inflexion near $B-V \approx 0$ (this corresponds to spectral type A) is associated with high absorption of ultraviolet radiation by the hydrogen in these stars.

Table 1 lists some data on the colors $B-V$ and $U-B$ of main-sequence stars and red giants of various spectral types (Johnson and Morgan).

TABLE 1

Sp	$B-V$	$U-B$	Sp	$B-V$	$U-B$	Sp	$B-V$	$U-B$
				Main sequence stars				
B0 V	−0.32	−1.13	A3 V	+0.09	+0.07	G5 V	+0.68	+0.21
B1 V	−0.28	−1.00	A5 V	+0.15	+0.09	G8 V	+0.70	+0.24
B2 V	−0.24	−0.86	A7 V	+0.19	+0.08	K0 V	+0.82	+0.48
B3 V	−0.20	−0.71	F0 V	+0.30	+0.02	K1 V	+0.86	+0.54
B5 V	−0.16	−0.56	F2 V	+0.37	+0.00	K3 V	+1.01	+0.89
B7 V	−0.13	−0.47	F5 V	+0.44	+0.00	K5 V	+1.18	+1.12
B8 V	−0.09	−0.29	F6 V	+0.47	−0.02	K7 V	+1.37	+1.26
B9 V	−0.05	−0.16	F8 V	+0.53	+0.02	M1 V	+1.48	+1.21
A0 V	0.00	0.00	G0 V	+0.60	+0.06	M3 V	+1.49	+1.10
A1 V	+0.05	+0.05	G2 V	+0.64	+0.16	M5 V	+1.69	+1.26
				Red giants				
G8 III	+0.95	+0.72	K2 III	+1.16	+1.20	K5 III	+1.52	+1.84
K0 III	+1.01	+0.86	K3 III	+1.30	+1.44	M2 III	+1.57	+1.86

§114. Mass—luminosity diagram. A close statistical correlation is observed between stellar luminosities and masses. The higher the mass, the higher the luminosity. The first empirical relation between these physical characteristics of stars was derived by Hertzsprung in 1919. Subsequently the mass—luminosity relation was studied by various authors.

Figure 156 is a diagram plotting the mass against the bolometric magnitude, obtained in 1937 by Parenago from the data on the masses of hot giants, including 53 visual binaries, 14 eclipsing binaries, and 6 average mass values for spectroscopic binaries. Note that the mass—luminosity relation is determined only for the components of binary systems, as it is only for these stars that the mass can be found. Figure 156 reveals a very close relation between the two parameters. If we assume that normal solitary stars form and evolve just like the binary stars do, the mass—luminosity relation defined by Figure 156 can be extended to normal stars as well. Estimates of the relative mass of giants in globular clusters, obtained by statistical methods, show that these stars on the whole follow the same dependence as the binary stars do. The mass and the luminosity of the Sun also satisfy the same relation.

The dependence shown in Figure 156 can be expressed by the empirical relation

$$\lg m = \underset{\pm 25}{0.590} - \underset{\pm 48}{0.1194}\, M_{\text{bol}}, \qquad (8)$$

where m is the stellar mass in units of solar mass. If the luminosity is also expressed in units of solar luminosity, relation (8) is found to be approximately equivalent to the equality

$$L = m^{3\frac{1}{3}}.$$

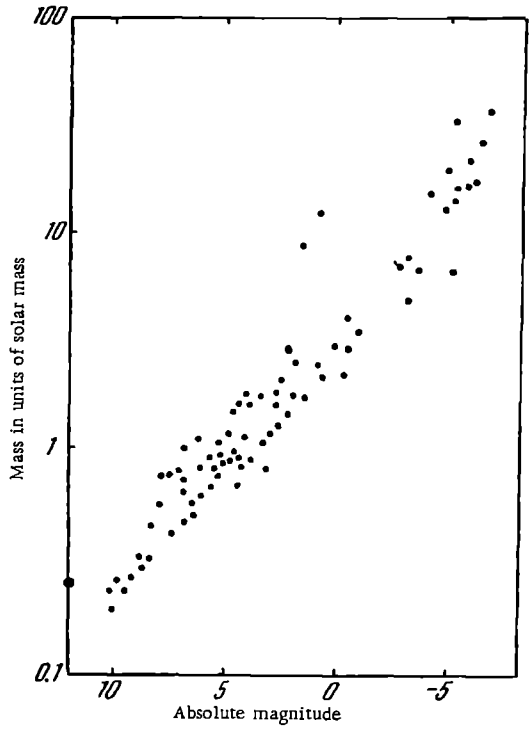


FIGURE 156

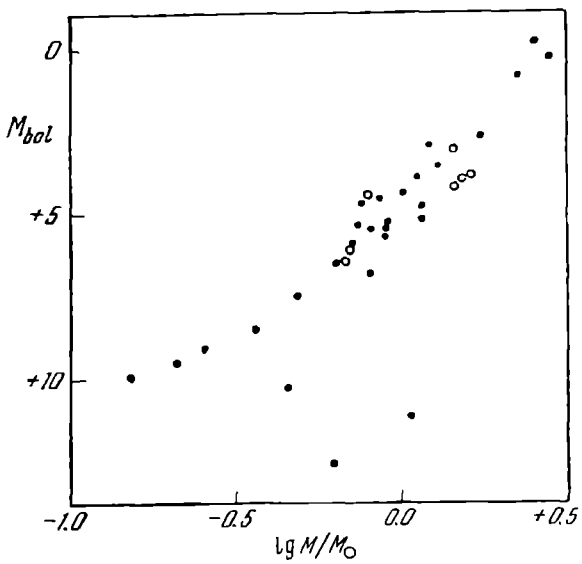


FIGURE 157

The interpretation of the mass—luminosity relation within the framework of the theory of stellar interiors is the subject of extensive literature.

Some star types markedly deviate from relation (8). This deviation is particularly pronounced for white dwarfs, whose mass is comparable with that of the Sun, whereas their luminosity is three orders of magnitude less.* Small systematic deviations of an opposite sign are observed for Algol variables. There are indications of other deviations too. In 1950 Parenago and Masevich came to the conclusion that each sequence in the spectrum — luminosity diagram has its own characteristic mass—luminosity relation.

For comparison we give here the mass—luminosity diagram (Figure 157) obtained in 1954 by Strand and Hall from the then available best data on the masses of binaries. In this diagram stars with error of less than 0.10 in the log mass are marked by dots, all the others are marked by circles. The three stars lying far from the common trend of all the other stars are white dwarfs.

§115. Luminosity function. Consider a certain galactic volume, small compared to the entire Galaxy but sufficiently large to contain numerous stars. Let $\Phi(M)$ be the fraction of stars in this volume with absolute magnitudes not higher than M. The function $\Phi(M)$ is called the integral luminosity function. The function $\varphi(M)=\Phi'(M)$ is called the differential luminosity function. Clearly $\varphi(M)dM$ is the fraction of stars in a given volume with absolute magnitudes between M and $M+dM$. By definition, the luminosity function is normalized to unity, so that $\lim_{M\to\infty}\Phi(M)=1$ and $\int_{-\infty}^{+\infty}\varphi(M)dM=1$.

This function is an important characteristic of the composition of the local stellar population in any given part of the galaxy. Since the composition of the stellar population is different in different parts of the Galaxy, the exact character of the luminosity function depends on the particular part of the Galaxy where this function is considered.

Knowledge of the luminosity function for each and every point in the Galaxy is clearly of the utmost importance. However, in very distant parts of the Galaxy, only giants and supergiants are visible. Meaningful determination of this function is therefore necessarily confined to some fairly small neighborhood of the Sun.

The main problem in the determination of the luminosity function is to minimize the observational selection. If, for example, we examine the list of all stars up to a given magnitude m, somehow estimate the distance r of each star, find the absolute stellar magnitudes using Eq. (6), and thus derive the frequency of the absolute magnitudes for this particular star sample, the result will be seriously distorted by observational selection. The fraction of giants will inevitably be greatly exaggerated, and the proportion of dwarfs too low. This is so because any list of stars up to a given magnitude contains giants from a very large volume of space and only those dwarfs which are nearest the Sun.

If we consider stars up to a certain distance r, the effect of observational selection is reduced, but nevertheless remains substantial, since the lists of the nearest stars are far from being complete, and numerous nearby dwarfs have not been discovered yet. The smaller the r we take, the more

* In 1958, V.V. Sobolev advanced a number of arguments implying that the currently accepted luminosities of the white dwarfs are too low.

complete are the lists, but they contain fewer stars and the sample may be not be sufficiently representative. In particular, this sample does not cover the rare giants and especially the supergiants, which are of the greatest significance for astrophysics and cosmogony.

TABLE 2

M_{ph}	$\varphi(M)$	M_{ph}	$\varphi(M)$
−6	0.00000013	8	0.031
−5	0.00000126	9	0.033
−4	0.0000048	10	0.047
−3	0.000019	11	0.069
−2	0.000060	12	0.100
−1	0.000126	13	0.107
0	0.00051	14	0.118
1	0.0023	15	0.118
2	0.0063	16	0.102
3	0.0078	17	0.079
4	0.017	18	0.049
5	0.024	19	0.020
6	0.033	20	0.0041
7	0.036	21	0.0008

To ensure a complete coverage of dwarfs and yet achieve a sufficiently large sample, Kapteyn, van Rhijn, Luyten and others proceeded from the statistical dependence between M, the proper motion μ, and the apparent magnitude m. This dependence is attributed to the fact that stars with large proper motions are inevitably nearby stars.

Soviet stellar astronomers generally use the luminosity function proposed by Parenago, which is a combination of van Rhijn's and Luyten's functions.

Table 2 lists the values of $\varphi(M)dM$ for $dM = 1$, and Figure 158 is the graph of this function.

The figure and the table show that the fraction of high-luminosity stars is vanishingly small. The mode of the curve corresponds to stars with $M = 14-15$, i.e., 10 magnitudes fainter than the Sun. The luminosity function is clearly bimodal, which implies that the entire statistical ensemble is divided into two subensembles — giants and dwarfs.

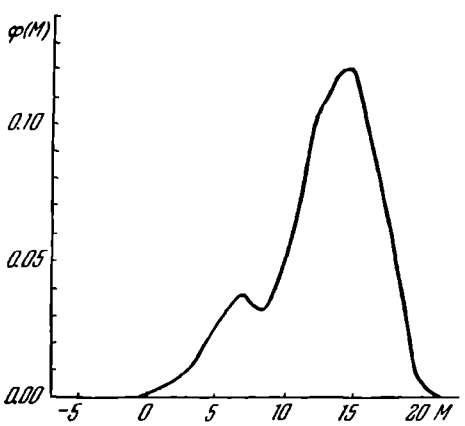


FIGURE 158

For absolute magnitudes over 15^m, the luminosity function is highly uncertain. According to Shatsova, Luyten's method underestimates the number of weak dwarfs, since the analysis inevitably skips the dwarfs with small tangential velocities. The mode of the function $\varphi(M)$ is thus possibly much higher that 15^m.

The fundamental parameters of the luminosity function are the following:

$$M_0 = 12.7 \qquad \mu_4 = 661$$
$$\mu_2 = 14.8 \qquad Sk = 0.711$$
$$\mu_3 = 40.5 \qquad \varepsilon = 0.017,$$

where M^0 is the mean value of the argument, $\mu_2 = \sigma^2$ is the variance, μ_3 and μ_4 are the central moments of third and fourth order, $Sk = \frac{\mu_3}{\mu_2^{3/2}}$ is the skewness, $\varepsilon = \frac{\mu_4}{\mu_2^2} - 3$ is the excess. The above data show that the luminosity function is very skewed. This means that the luminosity function is far from normal distribution.

The currently available data on the luminosity function are mainly confined to the neighborhood of the Sun, but they are of necessity used in the discussion of other parts of the Galaxy as well. This of course reduces the reliability of the results.

In addition to the general luminosity function, there are luminosity functions for different spectral types. For example, $\varphi_A(M)dM$ is equal to the ratio of stars of spectral type A with absolute magnitudes between M and $M+dM$ to the total number of stars. In this case $\int_{-\infty}^{+\infty} \varphi_A(M)dM$ gives the proportion of stars of spectral type A. This luminosity function is clearly normalized.

Since an almost exact single-valued relation is observed between the luminosity and the mass of a star, we can easily change over from the luminosity function to the distribution function of stellar masses. This gives the average stellar mass, which is found to be approximately 0.42 solar masses, or, approximately, $8.4 \cdot 10^{32}$ g.

§116. Apparent magnitude function. Let $N(m)$ be the number of stars observed in a certain part of the sky with apparent magnitudes not exceeding m. $N(m)$ is called the integral apparent magnitude function. It clearly depends on the size of the sky area used and the galactic coordinates of its center. The corresponding differential distribution function is $A(m) = N'(m)$.

Special functions $N_{Sp}(m)$ and $N_M(m)$ can be introduced. $N_B(m)$, for instance, gives the number of stars of spectral type B (in a given area of the sky) with apparent magnitudes not greater than m. $N_M(m)dM$ is the number of stars in a given sky area with apparent magnitude not exceeding m and absolute magnitude lying between M and $M+dM$. If the luminosity function is constant in this direction, we have

$$N_M(m)\,dM = N(m)\,\varphi(M)\,dM. \tag{9}$$

Similarly,

$$A_M(m)\,dM = N'_M(m)\,dM = A(m)\,\varphi(M)\,dM.$$

Seeliger's theorem. If the stars are uniformly distributed in space, the luminosity function constant, and the interstellar space perfectly transparent, then

$$\frac{N(m+x)}{N(m)} = 10^{0.6x}. \tag{10}$$

Indeed, if we take a fixed M and a small dM, $N_M(m)dM$ is equal to the number of stars of absolute magnitude between M and $M+dM$ inside a sphere of radius $r_{m,M}$ such that at this distance stars of absolute magintude M have apparent magnitude m. Since the luminosity function is constant throughout the volume, stars of this magnitude are distributed uniformly, so that

$$N_M(m)\,dM \sim r^3_{m,M}.$$

But as we know

$$H_m \sim r^{-2}_{m,M},$$

where H_m is the radiation flux from stars at distance $r_{m,M}$. Hence

$$N_M(m)\,dM \sim H_m^{-3/2}. \tag{11}$$

H_m can be found from Pogson's relation

$$m = -2.5 \lg H_m,$$

and insertion in (11) gives

$$N_M(m)\,dM \sim 10^{0.6m}. \tag{12}$$

By (12)

$$\frac{N_M(m+x)}{N_M(m)} = 10^{0.6x}. \tag{13}$$

For fixed M the theorem is thus clearly valid. If now $\varphi(M)$ is constant, Eq. (10) follows from (13) and (9), which completes our proof. Taking, in particular, $x = 1$, we get

$$\frac{N(m+1)}{N(m)} = 10^{0.6} \approx 3.98. \tag{14}$$

Differentiation of (12) with respect to m gives

$$A_M(m)\,dM \sim 10^{0.6m}.$$

For the function $A(m)$ we thus also have

$$\frac{A(m+x)}{A(m)} = 10^{0.6x}. \tag{15}$$

From Eqs. (14) and (15) we directly have

$$\lg N(m+1) - \lg N(m) = \lg A(m+1) - \lg A(m) = 0.6. \tag{16}$$

The validity of Eq. (16), and hence the applicability of the conditions assumed in Seeliger's theorem, is easily verified by counting stars to a given magnitude. Schiaparelli and Seeliger established that Eq. (16) was in fact not true. In any direction and for any m, the difference $\lg N(m+1) - \lg N(m)$ was found to be substantially less than 0.6.

This may be due to three factors: 1) thinning out of stars with distance from the Sun, 2) interstellar absorption of light, 3) peculiar luminosity function, such that far from the Sun the proportion of low-luminosity stars s greater than near the Sun.

Seeliger, Kapteyn, and others attributed the observed result

$$\lg N(m+1) - \lg N(m) < 0.6 \qquad (17)$$

entirely to the first factor. This implied that the Sun was situated at the point of maximum concentration of stars, i.e., at the center of the Galaxy. In fact, the main factor is interstellar absorption of light, as first noted by W. Struve (1847). In various directions in the galactic plane the stars do not thin out away from the Sun and inequality (17) is entirely associated with the effect of light absorption.

TABLE 3

m \ b	0°	10°	20°	30°	50°	70°	90°	$\frac{N(m,0)}{N(m,90)}$
4.0	8.193	8.117	7.989	7.870	7.744	7.689	7.655	3.44
	459	459	459	461	457	461	460	
5.0	8.652	8.576	8.448	8.331	8.201	8.150	8.115	3.44
	455	454	454	457	453	454	455	
6.0	9.107	9.030	8.902	8.788	8.654	8.604	8.570	3.44
	451	449	448	451	449	445	443	
7.0	9.558	9.479	9.350	9.239	9.103	9.049	9.013	3.51
	447	442	440	444	442	433	431	
8.0	0.005	9.921	9.790	9.683	9.545	9.482	9.444	3.64
	443	436	432	434	433	420	415	
9.0	0.448	0.357	0.222	0.117	9.978	9.902	9.859	3.88
	439	431	424	426	417	403	398	
10.0	0.887	0.788	0.646	0.543	0.395	0.305	0.257	4.27
	432	424	417	414	400	385	379	
11.0	1.319	1.212	1.063	0.957	0.795	0.690	0.636	4.82
	426	416	410	398	383	363	359	
12.0	1.745	1.628	1.473	1.355	1.178	1.053	0.995	5.62
	418	407	395	381	360	341	336	
13.0	2.163	2.035	1.868	1.736	1.538	1.394	1.331	6.79
	406	398	377	361	339	321	315	
14.0	2.569	2.433	2.245	2.097	1.876	1.715	1.646	8.38
	390	386	357	338	313	298	294	
15.0	2.959	2.819	2.602	2.435	2.189	2.013	1.940	10.45
	371	372	339	314	286	274	271	
16.0	3.330	3.191	2.941	2.749	2.475	2.287	2.211	13.15
	349	353	319	287	262	251	248	
17.0	3.679	3.544	3.260	3.036	2.737	2.538	2.459	16.60
	329	330	299	262	238	231	224	
18.0	4.008	3.874	3.559	3.298	2.975	2.769	2.683	21.13
	309	306	279	238	214	209	203	
19.0	4.317	4.180	3.838	3.536	3.189	2.978	2.886	26.98
	286	282	256	214	190	186	180	
20.0	4.603	4.462	4.094	3.750	3.379	3.164	3.066	34.43
	264	255	232	189	166	162	156	
21.0	4.867	4.717	4.326	3.939	3.545	3.326	3.222	44.16

In 1925 Seares, van Rhijn, Joyner, and Richmond published reduced counts of stars up to $18^m.5$. They used the listings of various catalogues for different parts of the sky or selected areas. All the stellar magnitudes were reduced to a single system, and the results were presented as values of $\lg N(m)$ averaged over 1 sq. degree at various galactic latitudes. The results are listed in condensed form in Table 3, which gives log star count up to a certain magnitude per degree for various m and b. The differences (in units of 10^{-3}) inscribed between the rows of the table give the values of

$$\lg N(m+1) - \lg N(m).$$

The table shows that these differences are substantially less than 0.6, rapidly decreasing with the increase in m. For small m, the differences $\lg N(m+1) - \lg N(m)$ are virtually constant for all b, but for higher m, $\lg N(m+1) - \lg N(m)$ in high galactic latitudes decreases much faster than in low latitudes. This is understandable since in low galactic latitudes the decrease of $\lg N(m+1) - \lg N(m)$ is primarily due to interstellar absorption of light, whereas in high latitudes the thinning out of stars away from the Sun also makes a substantial contribution. In this case, the variation of the luminosity function also plays a certain part, as the fraction of high-luminosity stars decreases away from the galactic plane.

The last column of Table 3 gives the ratio $\frac{N_0(m)}{N_{90}(m)}$ of the respective star counts per unit sky area at the galactic equator and the galactic pole. This ratio provides a numerical expression of the phenomenon of galactic concentration. The table shows that the galactic concentration rapidly increases with stellar magnitude.

The number of stars to given magnitude per unit sky area depends on the galactic longitude, as well as the galactic latitude. In 1928 Seares and Joyner published corrections to average $\lg N(m)$ values as a function of longitude for various m. The corrections for $m = 18$ are reproduced in condensed form in Table 4. The corrections are in units of 10^{-2} to the logarithm.

TABLE 4

l \ b	+70°	+50°	+30°	+10°	+0°	−10°	−30°	−50°	−70°
0°	+9	+20	+5	−30	−118	0	+25	+14	+6
20	+10	+7	+7	+16	−7	+10	+10	0	−1
40	+9	+10	0	+26	+25	+10	0	+13	−3
60	+4	+4	−3	+4	−2	+20	0	+8	−5
80	−1	+1	−10	−14	−20	−10	−10	−5	−6
100	−3	−6	−12	−24	−27	−12	−10	−16	−8
120	−5	−15	−20	−34	−36	−26	−26	−21	−5
140	−9	−17	−14	−20	−17	−25	−30	−16	+2
160	−13	−14	−12	−14	−27	−27	−29	−14	+2
180	−12	−12	−10	+10	+12	−25	−13	−10	−6
200	−6	−10	+2	+9	+11	−4	−16	−10	−16
220	0	0	−5	+4	+7	+3	−4	−11	−10
240	+3	+3	+2	+14	+18	0	−10	−12	−3
260	+6	+6	+4	+36	+70	+25	0	−10	−1
280	+6	+8	+9	+26	+36	+45	−2	0	0
300	+6	+10	+20	+25	+23	+50	+30	+14	+10
320	+6	+12	+8	+26	+42	+52	+40	+23	+11
340	+7	+14	+13	+5	+10	+38	+47	+20	+8

Table 4 shows that there is an excess of stars in directions close to 300–340° longitude (toward the galactic center) and a deficiency of stars in the directions toward the anticenter (100—160°). The longitudinal distribution of stars has a clearly pronounced simple wave.

The variation of corrections with longitude is clearly irregular. As the number of stars up to 18^m reaches thousands per 1 sq. degree, this irregularity cannot be explained by natural fluctuations in stellar distribution.

The contribution from various stellar clusters to this effect is insignificant. It is mainly associated with differences in the distribution of absorbing interstellar matter.

§117. The total number of stars in the Galaxy. Seares and van Rhijn have shown that for the range of m values where stellar counts are available the following empirical relation is satisfied:

$$\lg A(m) = \alpha + \beta m - \gamma m^2. \tag{18}$$

The coefficients α, β, and γ are clearly functions of latitude. For example, the mean variation of $\lg A(m)$ in galactic zones with b from 0 to 20°, from 20 to 40°, and from 40 to 90° gives the values of α, β, and γ listed in Table 5.

TABLE 5

b	α	β	γ	m_0	N
0° ± 20°	−5.520	+0.744	+0.0128	29.1	$2.84 \cdot 10^{10}$
± 20° ± 40°	−5.185	+0.704	+0.0135	26.1	$0.12 \cdot 10^{10}$
± 40° ± 90°	−5.303	+0.706	+0.0147	24.0	$0.02 \cdot 10^{10}$
				Total number of stars	$\approx 3 \cdot 10^{10}$

The deviations of the average values of $\lg A(m)$ from the results of Eq.(18) are fairly small. It was thus suggested that the apparent magnitude function could be written in the form

$$A(m) = 10^{\alpha+\beta m-\gamma m^2} \tag{19}$$

extrapolated over the entire range. We can thus estimate the total number of stars in a given direction. Indeed, since the coefficient γ is positive, (19) is a normal function. Simple algebraic manipulations will express it in terms of the standard parameters, i.e., the volume n, the mean value of the argument m_0, and the variance δ^2:

$$A(m) = 10^{\alpha+\beta m-\gamma m^2} = 10^{\alpha+\frac{\beta^2}{4\gamma}} e^{-\left(m-\frac{\beta}{2\gamma}\right)^2 \gamma \ln 10} = \frac{n_A}{\sigma\sqrt{2\pi}} e^{-\frac{(m-m_0)^2}{2\sigma^2}}$$

Equating

$$\frac{\beta}{2\gamma} = m_0,$$

$$\gamma \ln 10 = \frac{1}{2\sigma^2},$$

$$10^{\alpha+\frac{\beta^2}{4\gamma}} = \frac{n}{\sigma\sqrt{2\pi}},$$

TABLE 6

No.	Star	α δ	m	Sp	Proper motion, μ	Position angle of proper motion, θ	Trigonometric parallax π_{tr}, in thousandths of sec	Spectroscopic parallax π_{Sp}, in thousandths of sec	Radial velocity v_r, km/sec	M
1	+43°44 A	$0^h12^m.7$ +43°27	8.1	M2	2".90	82°	284± 5	250	+ 8±2	10.4
2	+43°44 B	. . . . .	10.9	M5	. . .	. . .	. . . .	. . .	. . . .	13.2
3	η Cas A	0 43 .0 +57 17	3.60	G0	1 .22	115	182± 5	143	+ 9±0	4.9
4	η Cas B	. . . . .	7.34	K6	. . .	. . .	. . . .		+ 13±2	8.6
5	van Maanen	0 43 .9 ± 4 55	12.4	wG	3 .00	155	243± 6	. . .	(+238)	14.3
6	τ Cet	1 39 .4 −16 28	3.66	G5	1 .92	297	298± 6	210	− 16±0	6.0
7	ε Eri	3 28 .2 − 9 48	3.75	K2	0 .98	271	303± 6	350	+ 15±0	6.2
8	40 Eri A	4 10 .7 − 7 49	4.52	K1	4 .08	213	202± 3	220	− 42±0	6.0
9	40 Eri B		9.62	wA	. . .	. . .	. . . .	. . .	. . . .	11.1
10	40 Eri C	. . . . .	11.1	M6	. . .	. . .	. . . .	. . .	. . . .	12.6
11	−45°1841	5 07 .7 −44 59	8.8	M0	8 .72	131	262± 6	. . .	+242±1	10.9
12	−3°1123	5 26 .4 − 3 42	8.1	M1	2 .24	160	166± 4	200	+ 11±1	9.2
13	Ross 42	5 26 .7 + 9 45	11.4	M6	0 .30	243	. . . .	250	. . . .	13.4
14	Ross 47	5 36 .4 +12 29	11.7	M5	2 .54	127	166± 8	170	. . . .	12.8
15	−21°1377	6 06 .4 −21 49	8.2	M0	0 .72	191	183± 6	150	. . . .	9.5
16	Ross 614	6 24 .3 − 2 44	11.3	M6	1, 01	135	260± 5	360	. . . .	13.4
17	α CMaA	6 40 .7 −16 35	−1.52	A1	1 .32	204	376± 3	360	− 8±0	1.4
18	α CMaB	. . . . .	8.54	wA5	. . .	. . .	. . . .	. . .	(+12)	11.4
19	−44°3045	6 54 .9 −44 09	10.8	M5	1 .13	264	. . . .	170	. . . .	12.3
20	+5°1668	7 22 .0 + 5 32	10.0	M4	3 .75	171	262± 6	240	+ 22±2	12.1
21	α CMi A	7 34 .1 + 5 29	0.53	F4	1 .25	214	291± 4	400	− 3±0	2.8
22	α CMi B	. . . . .	10.8	. . .	. . .	. . .	. . . .	. . .	. . . .	13.1
23	L 745—46	7 35 .8 −17 10	13.1	wF	1 .29	117	177±14	. . .	. . . .	14.3
24	Ross 882	7 39 .4 + 3 48	11.6	M6	0 .64	218	. . . .	230	. . . .	13.4

25	L 97—12	7 52 .6	−67 30	14.5	F—G	2 .05	135	170±30	. . .	. . . .	15.6
26	+50°1725	10 05 .3	+49 58	6.54	K8	1 .45	249	220± 8	260	− 27±1	8.2
27	+20°2465	10h14m.2	+20°22	9.46	M5	0″.49	265	193± 8	330	+ 9±2	10.9
28	L 1113—55	10 30 .9	+ 5 38	12.2	M6	0 .68	280	. . . .	170	. . . .	13.4
29	Wolf 359	10 51 .6	+ 7 36	13.5	M8	4 .76	234	403±10	360	−90	16.5
30	+36°2147	10 57 .9	+36 38	7.47	M2	4 .78	187	388± 6	380	− 87±2	10.4
31	+44°2051 A	11 00 .5	+44 02	8.7	M0	4 .53	272	174± 8	140	+ 64±4	10.9
32	+44°2051 B	. . . . .		14.8	M8	. . .	. . .	. . . .	. . .	. . . .	16.0
33	AC 79°3888	11 41 .3	+79 14	11.0	M4	0 .87	57	201±10	150	. . . .	12.5
34	Wolf 424 A	12 28 .4	+ 9 34	12.7	M7	1 .84	278	225± 7	250	. . . .	14.5
35	Wolf 424 B	. . . . .		12.7		. . .	. . .	. . . .	. . .	. . . .	. . .
36	+15°2620	13 40 .7	+15 26	8.58	M1	2 .30	129	191± 8	+180	+ 15±1	10.0
37	Prox Cen	14 22 .8	−62 15	11.3	M	3 .85	282	762± 5	. . .	. . . .	15.7
38	α Cen A	14 32 .8	−60 25	0.33	G4	3 .67	281	756± 7	. . .	− 22±0	4.7
39	α Cen B	. . . . .		1.70	K1	. . .	. . .	. . . .	. . .	. . . .	6.1
40	HR 5568 A	14 51 .6	−20 58	5.90	K4	1 .99	149	172± 4	180	+ 20±0	7.1
41	HR 5568 B	. . . . .		8.08	M0	. . .	. . .	. . . .	. . .	+ 25±4	9.2
42	−40°9712	15 25 .7	−40 54	10.2	M4	1 .57	229	166± 5	170	. . . .	11.3
43	−12°4523	16 24 .7	−12 25	9.8	M4	1 .17	186	255± 5	210	. . . .	11.8
44	36 Oph A	17 09 .2	−26 27	5.17	K2	1 .24	203	179± 5	180	− 1±0	6.4
45	36 Oph B	. . . . .		5.20	K1			. . . .	. . .	− 1±0	6.5
46	36 Oph C	. . . . .		6.53	K6	. . .	. . .	. . . .	. . .	− 10+1	7.8
47	−46°11540	17 21 .1	−46°47	9.7	M3	1 .07	148	225± 6	180	. . . .	11.5
48	−44°11909	17 29 .8	−44 14	10.9	M5	1 .15	217	208± 6	190	. . . .	12.5
49	+68°946	17 37 .0	+68 26	9.13	M3	1 .31	196	213± 5	240	− 17±3	10.8
50	+ 4°3561	17 52 .9	+ 4 25	9.46	M5	10 .25	356	543± 3	440	−110±4	13.1
51	70 Oph A	18 00 .4	+ 2 31	4.19	K1	1 .13	167	196± 4	230	− 7±0	5.7
52	70 Oph B	. . . . .		5.87	K5	. . .	. . .	. . . .	. . .	− 7±0	7.3
53	+59°1915 A	18 41 .7	+59 29	8.92	M3	2 .28	324	282± 4	240	0±2	11.2
54	+59°1915 B	. . . . .		9.73	M4	. . .	. . .	. . . .	. . .	7±3	12.0
55	Ross 154	18 43 .6	−23 56	10.5	M5	0 .67	106	350± 6	200	. . . .	13.2
56	+4°4048	19 12 .1	+ 5 02	9.2	M3	1 .46	204	170± 4	200	. . . .	10.4
57	L 347—14	19 13 .4	−45 42	13.4	M	2 .94	168	182± 9	. . .	. . . .	(14.7)
58	σ Dra	19 32 .6	+69 29	4.73	G9	1 .84	162	181± 4	170	+ 27±0	6.0

TABLE 6 (cont.)

No.	Star	α δ	m	Sp	Proper motion, μ	Position angle of proper motion, θ	Trigonometric parallax π_{tr}, in thousandths of sec	Spectroscopic parallax π_{Sp}, in thousandths of sec	Radial velocity v_r, km/sec	M
59	α Aql	$19^h45^m.9$ + 8 36	0.91	A5	0″.66	54	208± 5	170	− 26±2	2.5
60	δ Pav	19 58 .9 −66 26	67	G7	1 .65	134	174± 8	. . .	− 22±0	4.9
61	HR 7703 A	20 04 .6 −36°21	5.24	K2	1 .63	164	177± 9	170	−131± 0	6.5
62	HR 7703 B	. . . . .	11.5	M5	. . .	. . .	. . . .	. . .	. . . .	12.7
63	−32°16135 A	20 35 .6 −32 47	10.9	M5	0 .45	140	. . . .	200	+ 5	12.4
64	−31°17815	20 39 .0 −31 42	8.7	M1	0 .45	140	. . . .	170	+10	9.9
65	61 Cyg A	21 02 .4 +38 15	5.35	K3	5 .20	52	296± 3	200	− 65± 0	7.7
66	61 Cyg B	. . . . .	6.05	K5	. . .	. . .	. . . .	. . .	− 63± 1	8.4
67	−39°14192	21 11 .4 −39 15	6.65	M0	3 .47	251	257± 7	310	+ 22± 2	8.7
68	−49°13515	21 26 .9 −49 26	8.9	M3	0 .82	185	207±10	230	. . . .	10.5
69	ε Ind	21 55 .8 −57 12	4.70	K5	4 .69	123	288± 6	. . .	− 40± 0	7.0
70	−5°5715	22 04 .4 − 5 08	10.2	M4	1 .03	91	. . . .	170	. . . .	11.4
71	+56°2783 A	22 24 .4 +57 12	9.87	M4	0 .90	245	256± 4	220	− 24± 2	11.9
72	+56°2783 B	. . . . .	11.3	M6	. . .	. . .	. . . .	. . .	. . . .	13.3
73	L 789−6	22 33 .0 −15 52	12.3	M6	3 .27	46	328±11	230	. . . .	14.9
74	+43°4305	22 42 .5 +43 49	10.25	M5	0 .85	237	207± 7	250	+ 2± 2	11.8
75	−15°6290	22 47 .9 −14 47	10.27	M5	1 .11	123	220±11	250	. . . .	12.0
76	−36°15693	22 59 .4 −36 26	7.23	M0	6 .90	79	278± 6	270	+ 10± 1	9.4
77	Ross 248	23 37 .0 +43 39	12.2	M6	1 .82	176	314± 7	240	. . . .	14.7
78	+1°4774	23 44 .0 + 1 52	9.10	M2	1 .37	134	167± 6	170	− 64± 3	10.2
79	−37°15492	23 59 .5 −37 51	8.57	M3	6 .11	112	222± 7	260	+ 24± 2	10.3

we express the normal distribution parameters in terms of the coefficients α, β, and γ:

$$\left.\begin{aligned} m_0 &= \frac{\beta}{2\gamma}, \\ \sigma &= \frac{1}{\sqrt{2\gamma \ln 10}}, \\ n &= \sqrt{\frac{\pi}{\gamma \ln 10}} \cdot 10^{\alpha+\frac{\beta}{4\gamma}} \end{aligned}\right\} \qquad (20)$$

Eqs. (20) give the average apparent magnitude m_0 for all the stars in a given direction and the total number n of stars in 1 sq. degree. Multiplying n by the number of square degrees in a given zone, we obtain the total number N of stars in the zone. The calculated values of m_0 and N are listed in Table 5. The result shows that the bulk of stars are concentrated in the zone adjoining the galactic plane. The total number of stars in the Galaxy according to Seares and van Rhijn's estimates is $3 \cdot 10^{10}$.

§118. Star density function. Determination of star density near the Sun.

Let $D(r)$ be the number of stars in unit volume in a given direction at a distance r. $D(r)$ is called the star density function and its determination is one of the main problems of stellar statistics. If $D(r)$ were known in all directions, we would have all the main features of galactic structure.

Of particular significance is the determination of the star density near the Sun, i.e., $D(0)$, where it can be found with maximum reliability. This determination should naturally be based on lists of the nearest stars. In 1942 Kuiper published a list of stars with parallaxes $\geqslant 0''.095$, i.e., $r \leqslant 10.5$ pc. In 1957, Gliese extended the list of the nearest stars and included all the known objects with distances up to 20 pc. There are 915 objects in this list. Table 6 gives stars from Kuiper's list with $\pi \geqslant 0.166$, i.e., $r \leqslant 6$ pc. For comparison, Table 7 lists 25 of the brightest stars. Comparison of the two lists shows that of the 25 brightest stars only 4 are among the 79 nearest stars. This is associated with the great scatter of stellar luminosities — the nearest dwarfs are weaker than the far giants.

Within a sphere of 10.5 pc the stellar density can be regarded as constant, subject to natural fluctuations only. The density, however, cannot be found by simple division of the star count by the corresponding volume, since even the list of the nearest stars is definitely incomplete; many of the absolutely faint stars have not been discovered even in the immediate neighborhood of the Sun. To correct for this incompleteness, we divide the sphere of radius $r = 10.5$ pc into four equal concentric spherical volumes and using Kuiper's list count the stars in each of these subvolumes. If both the trigonometric and the spectroscopic parallax are known, the former should be used as it is more accurate. The results are listed in Table 8. Here r is the outer radius confining this volume, r' is the radius of the additional concentric spheres bisecting each volume into two equal parts, n is the number of stars in each volume, D^* is the mean density of known stars in each volume.

The table shows that D^* regularly decreases from near to far volumes, which is understandable, since with increasing distance the proportion of undiscovered stars naturally becomes higher. $D^*(r')$ can be represented by a linear equation

$$D^*(r') = a + br'. \qquad (21)$$

TABLE 7

No.	Star	α	δ	Sp	m_v	m_p	r_{nc}	M_v	Name
1	α CMa	6^h41^m	−16°35	A1V	−1.43	−1.43	2.66	+1.45	Sirius
2	α Car	6 22	−52 38	F0Ia	−0.73	−0.58	110	−6.0	Canopus
3	α Cen	14 33	−60 25	G2V	−0.27	+0.39	1.33	+4.1	
4	α Boo	14 11	+19 42	K2IIIp	−0.06	+1.17	11	−0.3	Arcturus
5	α Lyr	18 34	+38 41	A0V	+0.04	+0.04	8.1	+0.5	Vega
6	α Aur	5 09	+45 54	G5III+G0III	+0.09	+0.89	14	−0.7	Capella
7	β Ori	5 10	− 8 19	B8Ia	+0.15	+0.11	400	−8.2	Rigel
8	α CMi	7 34	+ 5 29	F5IV/V	+0.37	+0.78	3.48	+2.7	Procyon
9	α Eri	1 34	−57 45	B3V	+0.53	+0.37	24	−1.5	Achernar
10	β Cen	13 57	−59 53	B0V	+0.66	+0.45	63	−3.5	
11	α Ori	5 50	+ 7 23	M2Iab	+0.7	+2.6	150	−5.3	Betelgeuse
12	α Aql	19 46	+ 8 36	A7IV, V	+0.80	+1.02	5.0	+2.3	Altair
13	α Tau	4 30	+16 18	K5III	+0.85	+2.37	20	−0.6	Aldebaran
14	α Cru	12 21	−62 33	B0V	+0.87	+0.63	90	−4.0	
15	α Sco	16 23	−26 13	M1Ib	+0.98	+2.78	115	−5.2	Antares
16	α Vir	13 20	−10 38	B1V	+1.00	+0.77	63	−3.1	Spica
17	α PsA	22 52	−30 09	A3V	+1.16	+1.25	6.9	+1.95	Fomalhaut
18	β Gem	7 39	+28 16	K0III	+1.16	+2.17	11	+1.0	Pollux
19	α Cyg	20 38	+44 55	A2Ia	+1.26	+1.35	290	−6.2	Deneb
20	β Cru	12 42	−59 09	B0IV	+1.31	+1.08	115	−4.1	
21	α Leo	10 03	+12 27	B7V	+1.36	+1.25	21	−0.3	Regulus
22	ε CMa	6 55	−28 50	B2II	+1.49	+1.32	140	−4.5	
23	α Gem	7 28	+32 06	A1V	+1.59	+1.63	14	+0.9	Castor
24	λ Sco	17 27	−37 02	B2IV	+1.62	+1.39	80	−2.9	
25	γ Ori	5 20	+ 6 16	B2III	+1.64	+1.41	100	−3.4	

The data of Table 8 give a system of four equations for a and b. Solving this system by the least-squares method, we get

$$a = 0.120 \pm 0.008; \quad b = -0.086 \pm 0.010.$$

TABLE 8

No.	r	r'	n	D^*
1	6.63	5.26	93	0.0763
2	8.35	7.59	66	0.0541
3	9.56	9.00	48	0.0394
4	10.53	10.07	45	0.0369

The small mean errors indicate that Eq. (21) is valid and the numerical values of a and b are reliable; b characterizes the rate at which the incompleteness of our data grows with increasing distance. When the distance goes to zero, the data are complete so that $D^*(0)=D(0)$. But $D^*(0)=a$, and hence the star density near the Sun

$$D(0) = 0.120 \pm 0.008 \text{ stars per cubic parsec.}$$

In 1920 Kapteyn and van Rhijn estimated the star density near the Sun at 0.045 stars per pc^3, in 1939 Luyten found 0.103 stars per pc^3. Both these determinations are based on the statistical method of average parallaxes derived from the mean motions. Comparison of these figures with the above result, 0.12 stars per pc^3, shows that the later determinations give a higher

density. This is understandable, since both the list of stars with large proper motions and the list of the nearest stars are constantly revised and enlarged. It is readily seen, however, that if the luminosity function indeed has a peak at $M = 14-15^m$, the discovery of new stars should not affect the value of $D(0)$ obtained by extrapolation to zero. The situation will be different, however, if $\varphi(M)$ has a maximum at substantially higher M. The figure $D(0) = 0.12$ stars/pc^3 is then apparently too low.

§119. Fundamental equations of stellar statistics. Consider the stars observed within a solid angle ω. The volume element between spheres of radii r and $r+dr$ (Figure 159) is $\omega r^2 dr$, and the number of stars with apparent magnitudes between m and $m+dm$ in this volume is

$$\omega r^2 dr \cdot D(r) \cdot \varphi(M)\, dM, \tag{22}$$

where M is given by the equality

$$M = m - 5 \lg r + 5,$$

so that $dM = dm$.

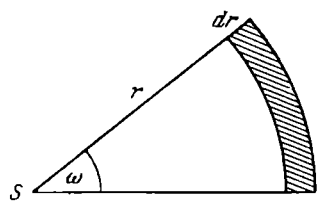


FIGURE 159

To find the total number of stars in a solid angle ω with apparent magnitudes between m and $m+dm$, (22) should be integrated over all r. Cancelling dm and dM, we get

$$A(m) = \omega \int_0^\infty D(r)\, \varphi(M)\, r^2\, dr. \tag{23}$$

It is assumed that $\varphi(M)$ is constant for all r. Eq. (23) is the fundamental equation of stellar statistics. It relates the apparent magnitude function to the luminosity function and the star density function. Since $D(r)$ is generally the unknown function, Eq. (23) is an integral equation. If $\varphi(M)$ is also unknown, we require another equation for these functions.

To this end, note that the parallax of a star inside the volume element $\omega r^2\, dr$ is $\frac{1}{r}$. To find the average parallax of stars with apparent magnitudes between m and $m+dm$ within the solid angle ω, the parallax $\frac{1}{r}$ should be multiplied by the number of stars having the given parallax and the given apparent magnitude, i.e., multiplied by (22), integrated over r and then divided by the total number of stars of given apparent magnitude $A(m)dm$:

$$\pi(m) = \frac{\omega\, dM}{A(m)\, dm} \int_0^\infty D(r)\, \varphi(M)\, r\, dr. \tag{24}$$

Cancelling dm and dM and multiplying by $A(m)$, we get

$$\bar{\pi}(m)\cdot A(m)=\omega\int_0^\infty D(r)\,\varphi(M)\,r\,dr. \tag{25}$$

This is the second equation of stellar statistics. Eqs. (23) and (25) were derived by Schwarzschild in 1910, and are called after him. Let us reduce these equations to canonical form. Substituting a new variable

$$\varrho=5\lg r-5, \tag{26}$$

so that

$$M=m-\varrho. \tag{27}$$

we write in the integrals in (23) and (25)

$$\Delta(\varrho)=200\ln 10\cdot\omega\cdot 10^{0.6\varrho}\cdot D(10^{0.2\varrho+1}), \tag{28}$$

the result being Schwarzschild's equations in canonical form:

$$A(m)=\int_{-\infty}^{+\infty}\Delta(\varrho)\,\varphi(m-\varrho)\,d\varrho, \tag{29}$$

$$\bar{\pi}(m)\,A(m)=\int_{-\infty}^{+\infty}\Delta(\varrho)\cdot 10^{0.2\varrho-1}\,\varphi(m-\varrho)\,d\varrho. \tag{30}$$

Since M and ϱ are independent variables and

$$m=M+\varrho,$$

we can regard $\Delta(\varrho)$ as the distribution function of ϱ and formally write Eq. (29) as the standard rule of probability theory for finding the distribution function of the sum of two independent variables.

§120. General solution of the first equation of stellar statistics. Schwarzschild's exact solution of Eq. (29) is based on the following property of the Fourier transform.

Consider a function $f(x)$. Its Fourier transform is

$$f^*(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{ixy}f(x)\,dx. \tag{31}$$

The original function is similarly recovered by the inverse Fourier transformation

$$f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-ixy}f(y)\,dy. \tag{32}$$

Multiplying the two sides of (29) by $\frac{1}{\sqrt{2\pi}}e^{i\beta m}\,dm$ we integrate from $-\infty$ to $+\infty$:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{i\beta m}A(m)\,dm=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{i\beta m}\,dm\int_{-\infty}^{+\infty}\Delta(\varrho)\,\varphi(m-\varrho)\,d\varrho=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{i\beta\varrho}\Delta(\varrho)\,d\varrho\int_{-\infty}^{+\infty}e^{i\beta(m-\varrho)}\varphi(m-\varrho)\,d(m-\varrho).$$

Thus, from (31)

$$A^*(\beta) = \sqrt{2\pi}\Delta^*(\beta)\varphi^*(\beta)$$

and hence

$$\Delta^*(\beta) = \frac{1}{\sqrt{2\pi}} \frac{A^*(\beta)}{\varphi^*(\beta)}.$$

Using (32), we recover $\Delta(\varrho)$:

$$\Delta(\varrho) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\varrho\beta} \frac{A^*(\beta)}{\varphi^*(\beta)} d\beta, \tag{33}$$

which gives the solution of Eq. (29). Once $\Delta(\varrho)$ is known, the star density function can be obtained from the equality

$$D(r) = \frac{5}{\omega \ln 10} r^{-3} \Delta(5 \lg r - 5), \tag{34}$$

which follows from Eqs. (26) and (28).

§121. Approximate solution of the first equation of stellar statistics. The solution described in the previous section, despite all its theoretical advantages, is not practical as the calculations are much too complex in applications and require knowledge of the function $A(m)$ for all m.

In addition to the exact treatment, Schwarzschild proposed a solution based on the assumption that $A(m)$ and $\varphi(M)$ are normal distributions. He proved that in this case $\Delta(\varrho)$ is also a normal distribution. Eq. (29) readily gives the parameters of this distribution. This method, however, has two shortcomings: 1) it gives a very crude approximation, since the luminosity function is essentially not normal; 2) if $\Delta(\varrho)$ is a normal function, for $\varrho \to \infty$ it should go to zero as $e^{-\alpha\varrho^2}$, namely faster than $10^{0.6\varrho}$, so that by (34) $D(r) \to 0$ for $r \to 0$. Near the Sun, however, the star density is definitely not zero.

In 1936, Cramer proved the following strong theorem in probability theory: if $A(m)$ is a normal function, then both $\varphi(M)$ and $\Delta(\varrho)$ are normal functions.

Hence, as $\varphi(M)$ is not a normal function, $A(m)$ is not normal either. The tendency to represent the apparent magnitude function by a normal distribution over the entire range, which emerged from the work of Seares and van Rhijn, was thus erroneous. The estimate of the total number of stars in the Galaxy, based on extrapolation using a normal apparent magnitude distribution, is also wrong. If the skewness of the apparent magnitude function is taken into consideration in the extrapolation, the total number of stars in the Galaxy is found to be $1.5 \cdot 10^{11}$.

Seares, Brill, and Bok developed numerical methods for the solution of Eq. (23). Computers have recently been applied to solve this equation.

It should be noted, however, that the use of Eq. (23) has not led to the construction of reliable density functions for various directions in the Galaxy. This is so because 1) Eq. (23) is solved only approximately, 2) the luminosity function is known with insufficient reliability, 3) star counts in small solid angles cover only few stars, so that substantial chance deviations from the standard luminosity function are possible, 4) to solve the equation, we require the absorption function (see §125), which is either unknown altogether or is known very approximately.

The significance of Eq. (23) is mainly in its applicability to various theoretical manipulations for conversion between different distribution functions. For directions in the plane of the Galaxy, the density function can be treated as constant to first approximation even over relatively small distances, so that substitution (26) reduces Eq. (23) to the form

$$A(m) = 200 \lg 10 \cdot \omega D \int_{-\infty}^{+\infty} 10^{0.6\varrho} \varphi(m-\varrho)\, d\varrho, \tag{35}$$

which gives a fairly simple relation between the magnitude function and the luminosity function.

§122. **Light absorption.** A whole range of observational results indicate that light is absorbed in the interstellar space. The principal results are the following: a) the ratio $\frac{N(m+1)}{N(m)}$ is markedly less than 3.98 in all directions and it decreases with increasing m; b) far stars show a definite reddening compared to the nearest stars of the same spectral type; c) certain sky areas are highly deficient in stars, whereas nearby areas show a multitude of stars; d) if the interstellar space is assumed perfectly transparent and the apparent stellar magnitudes of stars in open clusters are used to estimate their distance moduli, which combined with the angular dimensions of the clusters give their linear size, we find that on the average the farther the cluster the bigger is its linear diameter. This ridiculous result does not obtain if interstellar absorption is introduced; e) the number of galaxies up to a given apparent magnitude per unit sky area rapidly decreases toward the plane of the Galaxy. Since we can hardly imagine that the galaxies are distributed symmetrically relative to the plane of our Galaxy, the immediate conclusion is that the observed distribution is the outcome of the presence of dark matter in the Galaxy; f) dark matter is also observed in other galaxies.

Let us consider the effect of interstellar absorption on the apparen stellar magnitude. Let H be the luminous flux of a star at some distance r from it. When the light traverses a path length dr, the flux is attenuated. The flux attenuation dH is produced by two factors. One of the factors is the increasing distance from the star. The corresponding term is

$$\frac{r^2}{(r+dr)^2} H - H \approx -2H\frac{dr}{r}.$$

Additional attenuation is caused by interstellar absorption. The corresponding term is proportional to the flux and the distance travelled:

$$-kH\,dr.$$

The coefficient k characterizes the ability of the medium to absorb light and is called the absorptivity, or the absorption coefficient per unit length of the medium. It depends on the properties of the absorbing particles and is proportional to the density of dark matter.

Thus,

$$dH = -2H\frac{dr}{r} - kH\,dr. \tag{36}$$

Dividing by H and integrating, we get

$$H = cr^{-2}e^{-kr}.$$

The arbitrary constant c is determined by the luminosity of the star. If $k = 0$, we obtain the luminous flux without absorption

$$H_0 = cr^{-2}.$$

Hence

$$H = H_0 e^{-kr}. \tag{37}$$

Here $\tau = kr$ is the optical thickness of the medium; $q = e^{-k}$ is the transmissivity of the medium, its transparency coefficient per unit length:

$$H = H_0 e^{-kr} = H_0 e^{-\tau} = H_0 q^r. \tag{38}$$

Taking the decimal logarithm of (37), multiplying by -2.5 and changing over to apparent magnitudes, we get

$$m = m_0 + 2.5 \lg e \cdot k \cdot r = m_0 + ar, \tag{39}$$

where

$$a = 2.5 \lg ek \cong 1.086\, k. \tag{40}$$

a (like k) is generally called the absorption coefficient. It is usually calculated for 1 kpc, so that if r is in parsecs, Eq. (39) takes the form

$$m - m_0 = \frac{ar}{1000}. \tag{41}$$

For constant a, the attenuation of stellar brightness is proportional to distance. The coefficient a thus shows by how many stellar magnitudes the apparent magnitude of a star increases due to absorption over a distance of 1000 pc.

Integrating (36), we regarded k as constant over the entire light path. In general, k is a function of r. In Eq. (38) then $\tau = \int_0^r k(r)\, dr$. Hence,

$$a(r) \approx 1.086\, k(r),$$

where $a(r)\, dr$ is the absorption, in apparent stellar magnitudes, over a path of length dr. The integral absorption, in apparent magnitudes, over the entire path is

$$G(r) = m - m_0 = \int_0^r a(r)\, dr. \tag{42}$$

$G(r)$ is called the absorption functions.

§123. Selective absorption. The absorption of light involves definite reddening of the star. This shows that the absorption is a function of wavelength. Blue light is more readily absorbed than red light. In this case the absorption coefficient a clearly depends on the particular system of apparent magnitudes used, whether photographic, visual, or other.

Let us write Eq. (42) for photographic and visual magnitudes:

$$G_p(r) = m_p - m_{0p} = \int_0^r a_p(r)\,dr,$$

$$G_v(r) = m_v - m_{0v} = \int_0^r a_v(r)\,dr.$$

Subtracting the bottom equation from the top and seeing that $m_p - m_v$ is the color index CI, we get

$$CI - CI_0 = \int_0^r [a_p(r) - a_v(r)]\,dr. \tag{43}$$

The left-hand side of (43), equal to the difference of the observed color index and the color index without absorption, characterizes the reddening. It is called the color excess CE of the star. The difference $a_p - a_v = c$ is the selective absorption coefficient. If we assume that interstellar absorption is everywhere caused by the same substance, a_p, a_v, and c are only functions of the density of that substance, being proportional to density. The ratio

$$\frac{a_p}{c} = \gamma_p \tag{44}$$

is then independent of the density of dark matter, being constant in all directions and over all distances.

The equality

$$CE\gamma_p = \int_0^r \gamma_p \cdot c(r)\,dr = \int_0^r a_p(r)\,dr = G_p(r) \tag{45}$$

which follows from (43) and (44) is highly significant for the determination of interstellar absorption. It shows that the absorption of a star's light is proportional to the color excess of the star multiplied by a constant factor γ_p. The color excess can be found by subtracting from the observed color index CI the standard color index CI_0 of the corresponding spectral type without absorption. The constant factor γ_p is known fairly reliably, as it has been determined by numerous authors. On the average, γ_p is equal to 5. An analog of Eq. (45) can be written for absorption in visual stellar magnitudes.

Anyhow, if the absorption coefficient is constant, Eq. (45) takes the form

$$CE\gamma_p = \frac{a_p \cdot r}{1000}.$$

The determination of light absorption from color excess is currently the best of the available methods. Its shortcoming is that the spectral type of weak stars cannot be determined with adequate reliability. Moreover, the stars of any given spectral type are apparently characterized by a certain, though small, spread of color indices. This should introduce an obvious error in the determination of interstellar absorption.

The analytical dependence of the absorption coefficient at a given wavelength is of considerable theoretical and practical importance. There is an

extensive literature on this subject. It has been firmly established that in the visual and the photographic parts of the spectrum, the absorption is inversely proportional to wavelength.

This dependence corresponds to the case of light absorption by solid dust particles with radii of the order of 10^{-5} cm.

§124. Interstellar absorption and relation of M to m. Consider a star whose light is free from interstellar absorption. For this star we have

$$m = M + 5 \lg r - 5. \tag{46}$$

Now suppose that the space between the observer and the star is filled with absorbing matter. M and r remain as before, but m changes. Eq. (46) thus remains valid only if we put in the left-hand side m_0 — the apparent magnitude without absorption. But $m_0 = m - G(r)$ (see (42)); therefore in the presence of interstellar absorption

$$m = M + 5 \lg r + G(r) - 5. \tag{47}$$

Eq. (47) takes the form (46) if we introduce the so-called apparent distance r', defined by

$$5 \lg r' = 5 \lg r + G(r). \tag{48}$$

In particular, for a homogeneous absorbing medium,

$$5 \lg r' = 5 \lg r + \frac{ar}{1000}. \tag{49}$$

Clearly, if interstellar absorption is ignored and the absolute and apparent stellar magnitudes are used to find the distance from (46), the result is the apparent distance r', and not the true distance r. Since $G(r)$ and a are positive, the apparent distance is always greater than the true distance. Neglect of absorption thus gives an overestimated result for the distance.

If interstellar absorption is known, the true distance is found from the apparent distance by solving Eq. (48), which is easily done by the method of successive approximations. Table 9 gives the true and apparent distances for various values of a.

TABLE 9

r \ a	0.5	1.0	1.5	2.0	2.5	3.0	3.5
100	102	105	107	110	112	115	118
200	209	219	230	241	252	264	277
300	321	344	369	395	423	454	486
400	439	481	527	578	634	695	762
500	561	630	706	793	889	998	1096
600	689	791	908	1042	1197	1374	1578
700	822	966	1135	1334	1567	1841	2163
800	961	1156	1390	1671	2009	2415	2904
900	1107	1361	1675	2061	2535	3119	3837
1000	1259	1585	1995	2512	3162	3981	5012

The table shows that the distance can be greatly exaggerated if the interstellar absorption is ignored. We further see that the relative exaggeration

is greater for the distant stars than for the near ones. This is responsible for the apparent decrease of stellar density with increasing distance and for the decrease in the ratio $\frac{N(m+1)}{N(m)}$.

§125. Interstellar absorption and the fundamental equation of stellar statistics. The argument of the magnitude function is the apparent stellar magnitude augmented by interstellar absorption. Therefore, arguing as before and using the absorption function, we obtain the equation

$$A(m)=\omega\int_0^\infty r^2D(r)\,\varphi(M)\,dr, \tag{50}$$

where

$$M=m-5\lg r-G(r)+5.$$

Eq. (50) can be written in the form

$$A(m)=\omega\int_0^\infty r^2D(r)\,\varphi[m-5\lg r-G(r)+5]\,dr. \tag{51}$$

Suppose that we neglect the interstellar absorption and solve the equation

$$A(m)=\omega\int_0^\infty r'^2D_1(r')\cdot\varphi[m-5\lg r'+5]\,dr' \tag{52}$$

(the variable of integration is differently designated). Clearly the $D_1(r')$ obtained by solving Eq. (52) is not the same as the $D(r)$ obtained from Eq. (51). We call $D_1(r')$ the apparent density function, to distinguish it from $D(r)$ which is the true density function.

Substituting $5\lg r+G(r)=5\lg r'$, we change over in (52) to a new variable of integration r. This gives

$$A(m)=\omega\int_0^\infty r^2\cdot 10^{0.6G(r)}[1+0.2\ln 10\cdot rG'(r)]\cdot D_1[r\cdot 10^{0.2G(r)}]\times$$
$$\times\varphi[m-5\lg r-G(r)+5]\,dr. \tag{53}$$

Comparison of (51) and (53) shows that these equations coincide if we take

$$D(r)=10^{0.6G(r)}[1+0.2\ln 10\cdot rG'(r)]\cdot D_1[r\cdot 10^{0.2G(r)}]. \tag{54}$$

Eq. (54) expresses the true density function in terms of the apparent density function.

Thus, by solving the fundamental equation of stellar statistics without interstellar absorption, we get the apparent density function $D_1(r')$. If the absorption function $G(r)$ is known, Eq. (54) gives the true density function. If, conversely, we know the true density function, Eq. (54) is a differential equation for $G(r)$. Star counts thus give one of the functions $D(r)$, $G(r)$, if the other is known.

§126. Mean absorption function. Investigation of interstellar absorption in various directions (e.g., from color excesses) shows that the absorption reaches its maximum in the plane of the Galaxy and falls off rapidly with increasing galactic latitude. Comparison of the color excesses of near and

far objects in high galactic latitudes shows that in these latitudes the absorption remains virtually constant above a certain distance. These findings indicate that the absorbing matter is concentrated in a relatively thin layer near the galactic plane. A similar conclusion emerges from inspection of the photographs of other galaxies, which reveal the presence of dark matter.

Dynamic considerations indicate that this thin dark layer is of necessity inhomogeneous and does not have a sharp boundary. If the Galaxy is regarded as a highly flattened system with a symmetry plane and concentration of mass toward this plane, the distribution density of any group of gravitating objects should be a rapidly decreasing function of z — the distance from the plane of the Galaxy. Dynamic considerations show that this distribution is represented to fair approximation by the barometric function.

If the density of dark matter is indeed described by the barometric function, we can easily find the mean absorption function for a given galactic latitude and a given distance. Indeed, since the absorption coefficient a is proportional to density, the barometric equation is also applicable to a. At a given galactic latitude b and distance r we thus have

$$a(r, b) = a_0 e^{-\frac{z}{\beta}} = a_0 e^{-\frac{r \sin b}{\beta}}. \tag{55}$$

Multiplying (55) by dr and integrating from zero to r, we obtain the absorption function

$$G(r, b) = \frac{a_0 \beta}{\sin b}\left(1 - e^{-\frac{r \sin b}{\beta}}\right). \tag{56}$$

In (55) a_0 stands for the value of the variable in the principal plane. From (56) we actually find

$$G(r, 0) = a_0 r.$$

Taking $r = \infty$, $b = 90°$, we get

$$G(\infty, 90°) = a_0 \beta,$$

which shows that β is the half-thickness of the homogeneous layer of dark matter whose density is equal to the density of dark matter in the plane of the Galaxy.

Eq. (56) was derived by Parenago, assuming distribution (55) and ignoring possible fluctuations in the distribution of dark matter. Therefore, if we are dealing with a certain individual object, the interstellar absorption is best found from its color excess or by some alternative method in the corresponding direction. Eq. (56) should be applied to individual objects only if no other absorption data are available; remember that it only gives the mean absorption, definitely not the true absorption.

The great value of Parenago's equation is in statistical work. In this case, when we are dealing with a large volume of data covering a substantial part of the sky, the significance of density fluctuations is negligible and Eq. (56) gives useful results.

In conclusion of this section we give a summary of a_0 and $a_0 \beta$ values determined by various authors over the years (Tables 10 and 11). The values of a_0 obtained by Parenago, Wilkens, and Florya are markedly greater than the results of other authors, who in fact determined the

absorption coefficient for some arbitrary range of galactic latitudes near the plane of the Galaxy and not for $b = 0$ exactly. The result of these determinations is thus not a_0 but some average a which depends on the range of galactic latitudes and on the distance of the objects used.

TABLE 10

Author	Year	a_0	Author	Year	a_0
Trumpler	1930	$0^m.67$	Wilkens	1945	$3^m.2$
Joy	1933	0 .85	McLaughlin	1945	1 .05
Bok	1939	1 .0	Florya	1947	3 .3
Eigenson	1939	1 .2	Deich	1947	1 .3
Trumpler	1940	0 .70	Vashakidze	1949	1 .0
Kukarkin	1940	2 .0	Badalyan	1949	2 .05
Camm	1944	1 .36	Agekyan	1949	1 .6
Parenago	1945	3 .5	Mel'nikov	1951	1 .1

TABLE 11

Author	Year	$a_0\beta$
Hubble	1934	$0^m.25$
Williams	1934	0 .305
Oort	1938	0 .31
Parenago	1940	0 .34
Markaryan	1946	0.23—0.40
Vashakidze	1953	0 .34
Agekyan	1955	0 .46

§127. **The Vashakidze—Oort method for the determination of star density.** Let us assume at this stage that the star density and the density of dark matter are only functions of the distance from the plane of the Galaxy, i.e.,

$$D = D(z), \qquad a = a(z).$$

We can now write the fundamental integral equation of stellar statistics (50) for a certain direction and galactic latitude b, using a new variable of integration $z = r \sin b$:

$$A_b(m) = \omega \operatorname{cosec}^3 b \int_0^\infty z^2 D(z) \varphi [m - 5 \lg z + 5 - 5 \lg \operatorname{cosec} b - G(z) \operatorname{cosec} b] \, dz, \qquad (57)$$

where $G(r) = \int_0^r a(r)\, dr = \int_0^z a(z) \operatorname{cosec} b \, dz = G(z) \operatorname{cosec} b$ is the absorption in a given direction to a distance z from the plane of the Galaxy, $G(z)$ being the absorption to distance z in the direction of the galactic pole. In Eq. (57), the argument m is replaced by $m + 5 \lg \operatorname{cosec} b + \bar{G}$:

$$A_b [m + 5 \lg \operatorname{cosec} b + \bar{G}] =$$

$$= \omega \operatorname{cosec}^3 b \int_0^\infty z^2 D(z) [m - 5 \lg z + 5 + \bar{G} - G(z) \operatorname{cosec} b] \, dz.$$

$\bar{G}$ is chosen so that

$$\int_0^\infty z^2 D(z) \varphi [m - 5 \lg z + 5 + \bar{G} - G(z) \operatorname{cosec} b] \, dz = \int_0^\infty z^2 D(z) \varphi (m - 5 \lg z + 5) \, dz.$$

Clearly $\bar{G}$ can be interpreted as some mean absorption to stars of magnitude $m + 5 \lg \operatorname{cosec} b$. We may therefore write

$$A_b [m + 5 \lg \operatorname{cosec} b + \bar{G}_b] = \omega \operatorname{cosec}^3 b \int_0^\infty z^2 D(z) \varphi (m - 5 \lg z + 5) \, dz. \qquad (58)$$

If $b = 90°$, Eq. (58) takes the form

$$A_{90}[m + \overline{G_{90}(m)}] = \omega \int_0^\infty z^2 D(z)\, \varphi(m - 5\lg z + 5)\, dz. \qquad (59)$$

Comparison of Eqs. (58) and (59) gives

$$\sin^3 b A_b [m + 5\lg \operatorname{cosec} b + \overline{G_b(m + \lg \operatorname{cosec} b)}] = A_{90}[m + \overline{G_{90}(m)}]. \qquad (60)$$

Both sides of (60) can be found from observation. Only the determination of absorption involves certain difficulties. If the expressions in the left- and right-hand sides of (60) are indeed found to be equal to each other, this provides an indication of the validity of the starting assumption, namely that the surfaces of equal star densities are planes parallel to the plane of the Galaxy.

If, however, the left-hand side of Eq. (60) shows a systematic deviation from $A_{90}[m + \overline{G_{90}(m)}]$ for various b, the deviations provide an indication of the curvature of the surfaces of equal star densities and thus present a picture of the spatial behavior of the density function.

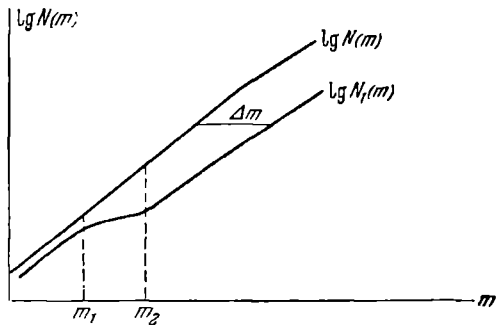
The Vashakidze—Oort method, as we see, should be applied to determine the general variation of the density function over large-scale volumes in the Galaxy. It is inapplicable to study the star density in a given direction. This method gives best results if applied to determine the special star density of objects of relatively high luminosity with a small spread in absolute magnitudes. The first property ensures penetration over large distances, which makes the method more effective. The second property establishes m as a reliable yardstick of distance; the surfaces of equal density in this case are determined with fair certainty.

For small b the left-hand side of (60) is uncertain because of the smallness of $\sin^3 b$ and the large value of the second factor. The Vashakidze—Oort method is generally limited to b above 25°.

§128. The significance of dark nebulae in interstellar absorption. In low galactic latitudes, which on the whole show high star densities, certain regions are anomalously deficient in stars. Originally these were regarded as actual voids in the Galaxy. Eventually it was established, however, that the observed deficiency of stars in certain direction is an apparent effect, associated with peculiar distribution of dark matter at large distances from the Sun: it obscures the light of more distant stars and actually renders many of them invisible. This conclusion emerged from the following arguments: a) because of differential galactic rotation, extended star-free corridors cannot persist in the Galaxy for a long time; b) the dark sky areas also show a low count of extragalactic nebulae; c) in the dark sky areas, the color excess of stars at distances of 160—260 pc, which lie beyond the dark clouds, is greater than in sky areas with normal star density. For more distant stars the difference in the color excess vanishes, since the dense sky areas may also have distant dark nebulae which cause reddening of the far stars but do not affect the total count of stars in the given direction.

M. Wolf proposed a method for studying the fundamental characteristics of dark nebulae based on comparison of the integral magnitude function in adjoining dark and normal sky areas. It is assumed that the observed differences in $N(m)$ are accounted for by an absorbing nebula in the dark area.

The plots of $\lg N(m)$ (for the light sky area) and $\lg N_1(m)$ (for the dark area) generally look like the curves in Figure 160.



[illegible]

For apparent magnitude m_1 the two curves almost coincide. It seems that the stars of these apparent magnitudes in the dark area lie on the near side of the nebula and the effect of the absorbing matter is therefore not felt. Starting with magnitudes m_1, the curves gradually diverge, since the growth of the function $\lg N_1(m)$ is slowed down. This indicates that the effect of the dark nebula has become noticeable. The nebula thus begins at the distance corresponding to the average distance of stars of apparent magnitude m_1. Since the average distance $\bar{r}(m_1)$ for stars of magnitude m_1 at a given galactic latitude is known,* we can find the distance where the nebula begins. We found $m = m_2$ the two curves have reached maximum separation, and after that they are practically parallel. The nebula thus ends at the distance $\bar{r}(m_2)$. The difference $\bar{r}(m_2) - \bar{r}(m_1)$ clearly gives the line-of-sight extent of the nebula. The horizontal distance between the parallel sections of the two curves gives the total absorption Δm in the dark nebula. Indeed, we have $\lg N_1(m) = \lg N(m - \Delta m)$, which is as it should be if the absorption in the nebula [illegible] the stars this side of the nebula should distort this regularity, but the near stars are too few to produce a noticeable effect).

The Wolf diagram can also be plotted laying $\lg A(m)$ off the vertical axis, where $A(m)$ is the number of stars with apparent magnitudes between $m - 1/2$ and $m + 1/2$. The argument in this case is precisely as before.

A particular development of Wolf's method calls for star counts in various spectral regions. An excellent example is provided by Franklin's study of the dark nebula in Cygnus. Star counts were made in photographic light up to $17^m.5$ ($P \leqslant 17.5$) and in red light up to $16^m.5$ ($R \leqslant 16.5$) in a normal area and in two adjoining dark areas. The diagrams are shown in Figure 161. They distinctly illustrate the effect of the dark nebula. From these graphs Franklin estimated the distance of the dark nebula (1600 pc) and obtained 0.[illegible] for the ratio of absorption in red light to absorption in photographic light. The diagram shows that in one of the dark areas the absorption in photographic light is $1.0-1^m.2$, and in the other it is $1.8-2^m.0$.

*The average distance of stars of a given magnitude is best determined by kinematic methods.

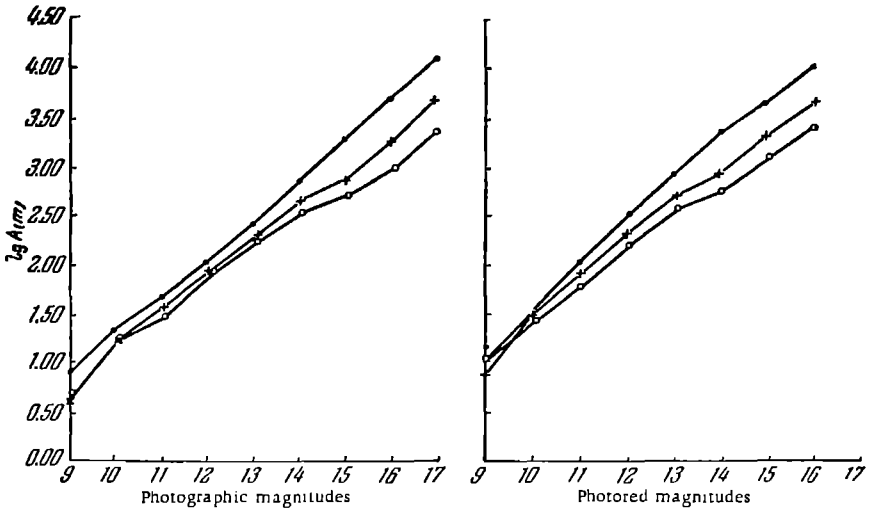


FIGURE 161

Application of the Wolf method has shown that it systematically gives too high results for the line-of-sight extent of the dark nebulae. The transverse linear dimensions of the nebulae, determined from their distance and the observed angular size, are generally one order of magnitude less than the line of sight extent. This result, if true, indicates that the dark nebulae are highly elongated formations aimed directly at the observer, which of course cannot be so.

Ogorodnikov has shown that a fundamental shortcoming of the Wolf method, which greatly distorts the characteristics of the nebulae, is the neglect of the spread in absolute stellar magnitudes. Indeed, the apparent magnitudes of stars inside the dark nebula need not all lie in the range between m_1 and m_2: other magnitudes are also possible. The section where the two curves diverge is therefore exaggerated, and this inevitably gives too high results for the line-of-sight extent of the nebula.

Ogorodnikov modified Wolf's method and took into consideration the variance of the luminosity function. This eliminated the systematic distortion. However, random fluctuations in the distribution of bright stars nevertheless make the improved Wolf method somewhat uncertain for the determination of distances and linear dimensions of individual dark nebulae.

The total absorption Δm obtained by the Wolf method, on the other hand, is highly reliable, since it is determined from the curve sections in the range of high stellar magnitudes, where the effect of random fluctuations is insignificant. The total absorption of dark nebulae determined in this way was found to be quite substantial, over 1^m and sometimes even over 2^m. It should be kept in mind, however, that only the darkest sky areas containing the bulkiest dark nebulae were investigated. These dark nebulae standing out prominently against the stellar background are known as "coal sacks."

Before the 1940's it was assumed that the dark matter filled a continuous region near the plane of the Galaxy and additionally collected in isolated, relatively few but extended and dense nebulae — coal sacks. Subsequently, however, our notions of the small-scale structure of dark matter have changed.

In 1938 Ambartsumyan and Gordeladze investigated the relation of bright diffuse nebulae to their exciting stars. Let R be the distance from the nebula to its exciting star, and L the luminosity of the star. If the nebula reflects the entire incident luminous flux, the energy per cm^2 of the star surface reaching the star in unit time from one square minute of the nebular surface as viewed from the star is

$$H = \frac{L}{4\pi (3438)^2} \frac{1}{4\pi R^2} . \tag{61}$$

Here 3438 is the number of angular minutes in one radian.

As we know, the surface brightness of extended objects is independent of distance, and therefore the quantity of energy per cm^2 of the Earth's surface reaching it from one square minute of the surface of the nebula as viewed from the Earth is also given by (61). Taking the logarithm of (61), multiplying by −2.5 and seeing that the absolute magnitude of a star is

$$M = -2.5 \lg \frac{L}{4\pi 10^2} ,$$

we find

$$m_n = M + 5 \lg R + 15.43, \tag{62}$$

where m_n is the apparent magnitude of 1 sq. minute of the nebula.

Bright nebulae were counted with the 60-in. Mount Wilson reflector with 1 hr exposures. This gave a limiting m_n of $14^m.36$. Substituting this value for m_n and solving Eq. (62) for R, we obtain for a given M the maximum distance of the exciting star from the nebula for which the nebula just shows in the photographs:

$$R_0 = 0.611 \cdot 10^{-0.2M}.$$

The illumination produced by the star is thus sufficient to make the diffuse nebulae visible within the volume

$$v(M) = \frac{4}{3} \pi R_0^3 = 0.955 \cdot 10^{-0.6M}. \tag{63}$$

To determine the fractional volume of the Galaxy which is sufficiently illuminated by stars of a given spectral type, we should multiply $v(M)$ by the number of stars of this spectral type with absolute magnitudes between M and $M+dM$ in 1 pc^3 and integrate over all M:

$$P_{Sp} = 0.955 D_{Sp} \int_{-\infty}^{+\infty} 10^{-0.6M} \varphi_{Sp}(M)\, dM. \tag{64}$$

Here $\varphi_{Sp}(M)$ is clearly a normalized function (see end of §143).

To calculate P_{Sp} from Eq. (64) we require D_{Sp} and φ_{Sp}, which however are insufficiently known. Therefore, in evaluating the right-hand side

of (64), we had better use the fundamental equation of stellar statistics written for stars of a given spectral type:

$$A_{Sp}(m) = \omega \int_0^\infty r^2 D_{Sp}(r)\, \varphi_{Sp}(M)\, dr. \tag{65}$$

If we count only the bright stars (which are on the average the closest to the Sun), we can first neglect interstellar absorption and second take the partial star density constant. In this case, changing over to a new variable of integration

$$M = m - 5 \lg r + 5,$$

we obtain

$$A_{Sp}(m) = 200 \ln 10 \omega D_{Sp} 10^{0.6m} \int_{-\infty}^{+\infty} 10^{-0.6M} \varphi_{Sp}(M)\, dM. \tag{66}$$

Dividing (64) through by (66), solving the resulting equation for P_{Sp}, and calculating the numerical coefficient, we get

$$P_{Sp} = 0.00207 \frac{1}{\omega} A_{Sp}(m)\, 10^{-0.6m}. \tag{67}$$

From Eq. (67) we readily find P_{Sp}. Using the stars with apparent magnitudes from 2.0 to 5.0 in Schlesinger's catalogue of bright stars, Ambartsumyan derived the P_{Sp} values of various spectral types (Table 12).

TABLE 12

Spectrum	P_{Sp}	Expected number	Actual number
O	0.00002	11	47
B0	0.00002	11	23
B1—B9	0.00022	120	122
A	0.00006	33	28
F	0.00005	27	6
G	0.000010	5	1
K	0.000045	25	5
M	0.000005	2	2
Total . . .	0.00043	234	234

The number of diffuse nebulae for which the spectra of the exciting stars have been determined is currently 234. The third colum of Table 12 gives the expected number of diffuse nebulae illuminated by stars of each spectral type, calculated assuming strict proportionality to the volume illuminated by stars of each spectral type. The fourth column gives the true number of nebulae illuminated by stars of each spectral type, according to Cederblad.

Comparison of third and fourth column reveals a definite correlation between the volume illuminated by stars of each spectral type and the number of diffuse nebulae illuminated by these stars. The assumption of random encounters between stars and diffuse nebulae is thus on the whole valid. On the other hand, the number of observed diffuse nebulae is clearly higher than the expected number of nebulae for stars of spectral types O and B0 and, conversely, there is a distinct deficiency of diffuse nebulae for stars of types F, G, and K. This effect can be due to two reasons. The first is that diffuse nebulae contain hydrogen, which typically degrades short-wave radiation into visible radiation. As a result, the brightness of diffuse nebulae illuminated by hot O and B0 stars, which are powerful sources of short-wave radiation,

is greatly enhanced, and the volume of space effectively illuminated by these stars is in fact greater than that calculated from Eq. (67). Moreover, recent investigations of hot giants and diffuse nebulae point to a possible genetic relationship between them. In this case, analysis of the tabulated data suggests that for stars of types B1—B9 and A the relationship with diffuse nebulae is relatively weak, though definitely felt, whereas for later type stars no relationship is observed altogether.

The sum of all P_{sp} shows that only 1/2300 of space is sufficiently illuminated to make the diffuse nebulae visible. Therefore, the established fact of random encounters between diffuse nebulae and stars shows that the number of unilluminated, i.e., dark, nebulae should be a factor of 2300 greater than the number of luminous nebulae. Photographs taken with 1-hr exposure through the 60-in. Mount Wilson reflector reveal a total of 250 luminous nebulae in a cylindrical layer of about 1500 pc radius with its axis directed through the Sun at right angles to the plane of the Galaxy. The same cylindrical layer should therefore contain $250 \times 2300 \approx 580{,}000$ dark nebulae. Since the radius of the Galaxy is approximately nine times the radius of this cylindrical layer, the total number of dark nebulae in the Galaxy is roughly $5.8 \cdot 10^5 \cdot 9^2 \approx 5 \cdot 10^7$. This figure should be reduced by a certain factor if we consider the fluorescence of gas in the nebulae and the presence of a genetic relationship between the nebulae and hot giants. Nevertheless, the number of dark nebulae in the Galaxy is over 10^7. This result first led Ambartsumyan to the conclusion that the absorption of light in the Galaxy is mainly caused by a multitude of individual dark nebulae. This important conclusion was subsequently confirmed by numerous investigators.

§129. Theory of field brightness fluctuations. The conclusion concerning the patchy, discrete structure of the absorbing intersteller medium was the first step in the development of Ambartsumyan's theory of fluctuations in the apparent distribution of galaxies and the Milky Way brightness. Markaryan, Rusakov, Agekyan, Chandrasekhar, Munch, and many others contributed to this theory. In the present section we will discuss in a somewhat more general form the fundamental concepts of Ambartsumyan's theory. As we have already noted, certain sky areas show distinct deficiency of stars. However, if we ignore these special areas and consider only normal sky areas at the same latitude, they also reveal marked fluctuations in the number of stars to any given magnitude. Thus, according to Markaryan, the relative mean square fluctuation in the number of stars to a given magnitude.

$$\sigma_N^2 = \frac{\overline{N^2(m)} - \overline{N(m)}^2}{\overline{N(m)}^2} \tag{68}$$

takes on different values at different galactic latitudes, as listed in Table 13.

TABLE 13

b	σ_N^2	σ_I^2	σ_{nat}^2
from $\pm$ 5 to 0°	0.116	0.115	0.001
from $\pm$ 8 to $\pm$11°	0.079	0.078	0.001
from $\pm$ 28° to $\pm$ 32°	0.027	0.023	0.004
from $\pm$ 38° to $\pm$ 43°	0.027	0.019	0.008

Under σ_I^2 the table gives the corresponding mean square fluctuation in the surface brightness of the stellar sky. If the surface brightness of the sky is entirely attributed to the individual stars, it can be found from the expression

$$I = c \sum_{m=-\infty}^{+\infty} \left[N\left(m+\tfrac{1}{2}\right) - N\left(m-\tfrac{1}{2}\right) \right] 10^{-0.4m} = c \sum_{m=-\infty}^{+\infty} A(m)\, 10^{-0.4m}. \qquad (69)$$

Here $c \cdot 10^{-0.4m}$, as we know, is the brightness of a star; c depends on the units of measurement and clearly drops out in the calculation of σ_I^2. Formally, to find I, we require star counts in the entire range of m values. However, as m increases by unity, $10^{-0.4m}$drops a factor of 2.5, whereas for $m \geqslant 12$, as we see from Table 3, $A(m)$ increases by less than a factor of 2.5 (even at 0° galactic latitude) and its rate if increase substantially slows down for higher magnitudes. At high galactic latitudes, $A(m)$ increases even at a slower rate. Therefore, stars up to 18^m (the limit in Table 12) determine most of the surface brightness of the stellar sky.

Table 12 shows that the mean square fluctuations in star count and surface birghtness are correlated (this is natural, since the surface brightness is determined by the number of stars) and increase at lower b.

The observed fluctuations may be due to the following reasons: a) natural fluctuations in the spatial distribution of stars; b) physical grouping of stars into clusters, associations, etc.; c) errors in stellar magnitude determinations; d) fluctuations in interstellar absorption.

The natural relative mean square fluctuation is given by

$$\sigma_{\text{nat}}^2 = \frac{1}{\overline{N(m)}}.$$

It will therefore decrease with decreasing b, since $\overline{N(m)}$ increases in this direction. This conclusion is clearly at variance with the observed trend. Moreover, the data under σ_{nat}^2 in the table show that the natural fluctuations in star counts are too small.

The clustering effect cannot be substantial, since the clusters incorporate a small minority of stars. The great majority of stars are the galactic field stars. The role of stellar associations is apparently more significant, since, though few in number, they contain high-luminosity stars. However, the number of stars in each association is small and they should therefore make a greater contribution to total sky brightness fluctuations than to star count fluctuations.

Markaryan's study has shown that fluctuations due to errors in the limiting magnitudes m are also negligible compared to the observed fluctuations. His analysis thus apparently confirms Ambartsumyan's result, namely that the fluctuation in the observed number of stars is mainly due to dark clouds.

Let us consider a general problem that can be applied to study fluctuations in the distribution of various objects.

Suppose a certain field is made up of luminous matter and absorbing (dark) matter. Both kinds of matter continuously fill the entire space according to a certain density distribution law and, furthermore, they collect into distinct agglomerations which are also distributed according to a certain law with natural fluctuations.

Consider an arbitrary direction S. The distance from the origin to some point A in this direction is s.

Let $\eta(s)$ be the radiation from a unit volume of the continuously distributed luminous matter in unit solid angle, $\tau(s)$ the optical thickness of the continuous dark matter per unit distance, $\varkappa(s)$ the probability of encountering a dark cloud per unit distance, $\lambda(s)$ the probability of encountering a bright cloud per unit distance, $\varphi(q)$ the normalized distribution function of the transparency coefficients of the dark clouds, ξ the radiation in unit solid angle from a cylinder of unit cross section with axis in the direction S in a bright cloud, $\chi(\xi)$ the corresponding normalized distribution function.

Let $\psi(x, s)$ be the probability to oberve from point A in the direction S a surface field brightness not exceeding x. Let us also consider a point B in the direction S, distant $s+\Delta s$ from the origin (Figure 162). Δs is regarded as a small increment.

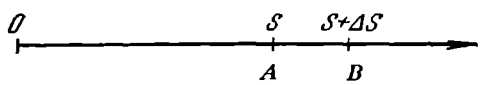


FIGURE 162

Let us establish the conditions at the point B and over the segment Δs such that the field brightness observed from the point A in the direction S does not exceed x. This will be so when one of the following three mutually exclusive events occurs.

The first event is that the brightness observed from the point B in the direction S is at most $x+x\tau\Delta s-\eta\Delta s$ (the corresponding probability being $\psi(x+x\tau\Delta s-\eta\Delta s, s+\Delta s)$), and the segment Δs is equally free from dark and luminous condensations. Then the continuously distributed dark matter over the segment Δs will reduce the radiation intensity by $x\tau\Delta s$, the continuously distributed luminous matter will increase the intensity by $\eta\Delta s$, so that the observed brightness at the point A will not exceed x. The probability that the segment Δs is free from any luminous and dark condensations is $(1-\varkappa\Delta s)(1-\lambda\Delta s)=1-\varkappa\Delta s-\lambda\Delta s+0(\Delta s)$, where $0(\Delta s)$ is of second order of smallness relative to Δs and is ignored in what follows. According to the multiplication theorem, the probability of the first event is therefore

$$\psi(x+x\tau\Delta s-\eta\Delta s,\ \ s+\Delta s)(1-\varkappa\Delta s-\lambda\Delta s). \tag{70}$$

The second event is that the brightness observed from B does not exceed x/q and the segment Δs is occupied by a dark cloud with a transparency coefficient q. The probability of this event is

$$\varkappa\Delta s\int_0^1\psi\left(\frac{x}{q},\ s+\Delta s\right)\varphi(q)dq, \tag{71}$$

where we multiplied by $\varphi(q)dq$ and integrated over q so as to account for dark clouds of all possible transparency coefficients. Note that in calculating the probability of the second event, we ignored the effect of the continuous distribution of luminous and dark matter. This is so because (71) is itself of the first order of smallness in Δs and the introduction of the increments

$x\tau\Delta s-\eta\Delta s$ in the argument of ψ to allow for the effect of the continuous distribution will make a second order contribution, which is ignorable.

The third event is that the brightness observed at the point B does not exceed $x-\xi$, and the segment Δs is occupied by a luminous condensation with emission ξ. The probability of this event, allowing for all possible ξ, is

$$\lambda\Delta s\int_{\xi}^{\infty}\psi(x-\xi,\ s+\Delta s)\,\chi(\xi)\,d\xi, \tag{72}$$

and the lower limit is chosen from the consideration that the brightness $x-\xi$ must not be negative.

The probability of obtaining a brightness not exceeding x at the point A with two or more condensations of any kind along Δs is ignored, since the probability of these events is of second and higher order of smallness in Δs.

Since the first, the second, and the third event are mutually exclusive, $\psi(x,\ s)$ according to the addition theorem is equal to the sum of the probabilities (70), (71), and (72):

$$\psi(x+x\tau\Delta s-\eta\Delta s,\ s+\Delta s)(1-\varkappa\Delta s-\lambda\Delta s)+$$
$$+\varkappa\Delta s\int_{0}^{1}\psi\left(\frac{x}{q},\ s+\Delta s\right)\varphi(q)\,dq+\lambda\Delta s\int_{\xi}^{\infty}\psi(x-\xi,\ s+\Delta s)\,\chi(\xi)\,d\xi=\psi(x,\ s). \tag{73}$$

Expanding $\psi(x+x\tau\Delta s-\eta\Delta s,\ s+\Delta s)$ in a Taylor series in powers of the first argument, we transfer the term $\psi(x,\ s)$ to the right-hand side of the equation, divide the two sides by Δs, and allow Δs to go to zero. In the limit, we get

$$\frac{\partial\psi(x,\ s)}{\partial s}-(\varkappa+\lambda)\psi(x,\ s)+(x\tau-\eta)\psi'_x(x,\ s)+$$
$$+\varkappa\int_{0}^{1}\psi\left(\frac{x}{q},\ s\right)\varphi(q)\,dq+\lambda\int_{\xi}^{\infty}\psi(x-\xi,\ s)\,\chi(\xi)\,d\xi=0. \tag{74}$$

Equation (74) is the fundamental equation of the theory of fluctuations.

Differentiating (74) with respect to x, we change over to the differential distribution function $f(x,\ s)=\psi'_x(x,\ s)$:

$$\frac{\partial f(x,\ s)}{\partial s}-(\varkappa+\lambda)f(x,\ s)+[(x\tau-\eta)f(x,\ s)]'_x+$$
$$+\varkappa\int_{0}^{1}f\left(\frac{x}{q},\ s\right)\frac{1}{q}\varphi(q)\,dq+\lambda\int_{\xi}^{\infty}f(x-\xi,\ s)\,\chi(\xi)\,d\xi=0. \tag{75}$$

Multiplying (75) by $x^n dx$ we integrate over x from zero to ∞. Let $\mu_n(s)$ be the initial moment of n-th order of the function $f(x,\ s)$:

$$\mu_n(s)=\int_{0}^{\infty}x^n f(x,\ s)\,dx. \tag{76}$$

Then, seeing that for $x\to\infty$, $f(x,\ s)$ should go to zero faster than any power of $1/x$, we get

$$\int_{0}^{\infty}x^n[(x\tau-\eta)f(x,\ s)]'\,dx=x^n(x\tau-\eta)f(x,\ s)\Big|_0^{\infty}-n\int_{0}^{\infty}x^{n-1}(x\tau-\eta)f(x,\ s)\,dx=-n\tau\mu_n+n\eta\mu_{n-1},$$

$$\int_0^\infty x^n\,dx\int_0^1 f\left(\frac{x}{q},s\right)\frac{1}{q}\;\varphi(q)\,dq=\int_0^1 q^n\varphi(q)\,dq\int_0^\infty\left(\frac{x}{q}\right)^n f\left(\frac{x}{q},s\right)d\,\frac{x}{q}=\overline{q^n}\mu_n(s),$$

$$\int_0^\infty x^n\,dx\int_\xi^\infty f(x-\xi,\,s)\,\chi(\xi)\,d\xi=\int_0^\infty\chi(\xi)\,d\xi\int_0^\infty(x+\xi)^n f(x,\,s)\,dx=$$

$$=\int_0^\infty\chi(\xi)\,d\xi\int_0^\infty\sum_{k=0}^{n}C_n^k\xi^k x^{n-k}f(x,\,s)\,dx=\sum_{k=0}^{n}C_n^k\overline{\xi^k}\mu_{n-k}=\mu_n+\sum_{k=1}^{n}C_n^k\overline{\xi^k}\mu_{n-k},$$

and Eq. (75) takes the form

$$\frac{d\mu_n}{ds}-[\varkappa(1-\overline{q^n})+n\tau]\,\mu_n+n\eta\mu_{n-1}+\lambda\sum_{k=1}^{n}C_n^k\overline{\xi^k}\mu_{n-k}=0. \tag{77}$$

Here $\overline{q^n}$ is the mean value of the transparency coefficient raised to the n-th power, and $\overline{\xi^k}$ is the mean value of ξ raised to the k-th power.

Eq. (77) is the general equation of the theory of fluctuations expressed in moments. This equation is useful for the solution of various particular problems.

§130. **Surface brightness fluctuations of the Galaxy.** To apply Eq. (77) to the distribution of the galactic component of the surface sky brightness, we should establish which of the two properties — continuity or discreteness — predominates in the structure of luminous and dark matter in the Galaxy. For the dark matter the corresponding estimate was derived in §124. It seems that the dark matter (dust) in the Galaxy occurs mainly in the form of isolated clouds, and the continuous substrate of dark matter is of negligible significance. In our case $\tau(s)=0$ and $\varkappa(s)>0$.

The luminous matter in the Galaxy are the stars. Although the stars are discrete objects, we must remember that the space between two neighboring dust clouds is on the average occupied by a very large number of stars. In this context stars should therefore be regarded as continuously distributed. If a substantial proportion of stars were found in stellar clusters, which occur at distances comparable with the distances between dark clouds, we would regard the luminous matter in the Galaxy as collected into discrete condensations. Only a small minority of stars occur in stellar clusters, and we may therefore take $\eta(s)>0$ and $\lambda(s)=0$.

For the surface brightness of the Galaxy, Eq. (77) is thus written in the simplified form

$$\frac{d\mu_n}{ds}-\varkappa(1-\overline{q^n})\,\mu_n+n\eta\mu_{n-1}=0. \tag{78}$$

Eq. (78) is a linear differential equation for $\mu_n(s)$. Its solution is

$$\mu_n=ne^{(1-\overline{q^n})\int\varkappa\,ds}\int_s^\infty\eta\mu_{n-1}e^{-(1-\overline{q^n})\int\varkappa\,ds}\,ds, \tag{79}$$

where the arbitrary constant is determined from the boundary condition $x=0$ for $s=\infty$, so that $\mu_n=0$.

Since $\mu_0=1$, the recursive formula (79) in principle can be used to calculate the moments μ_n for any $\eta(s)$ and $\varkappa(s)$.

Consider a particular case when $\eta(s)$ and $\varkappa(s)$ differ only by a constant factor h:

$$\eta(s) = h\varkappa(s). \tag{80}$$

If we remember that the bulk of the luminous matter in the Galaxy are stars, whose concentration toward the galactic plane is comparable with the concentration of the dark matter, we see that condition (80) is not far from the truth.

Writing (80) for $n = 1$ and seeing that $\mu_0 = 1$, we get

$$\mu_1(s) = he^{(1-\bar{q})\int_s^\infty \varkappa(s)\,ds} \int_s^\infty \varkappa(s)\, e^{-(1-\bar{q})\int_0^s \varkappa(s)\,ds}\, ds \tag{81}$$

and integration gives

$$\mu_1(s) = \frac{h}{1-\bar{q}}\left[1 - e^{-(1-\bar{q})\int_0^\infty \varkappa(s)\,ds}\right]. \tag{82}$$

Similarly, μ_2 is integrated by inserting (82) in (79), and so on for any n. By induction, we get

$$\mu_n(s) = n!\,h^n \sum_{i=0}^{n} \frac{\bar{q}^i}{\prod_{j=0}^{i}(\bar{q}^i - \bar{q}^j)}\, e^{-(1-\bar{q}^i)\int_0^s \varkappa(s)\,ds}. \tag{83}$$

We are interested in the case when the observer is situated at the point $s = 0$. Thus,

$$\mu_n(0) = n!\,h^n \sum_{i=0}^{n} \frac{\bar{q}^i}{\prod_{j=0}^{i}(\bar{q}^i - \bar{q}^j)}\, e^{-(1-\bar{q}^i)\int_0^\infty \varkappa(s)\,ds}. \tag{84}$$

If $\varkappa$ is only a function of the distance from the plane of the Galaxy, we have irrespective of the particular form of this function

$$\int_0^\infty \varkappa(s)\,ds = \zeta \operatorname{cosec} b, \tag{85}$$

where ζ is the expectation value of the number of dark clouds in the direction to the galactic pole, b is the galactic latitude.

In this case, for instance, Eq. (84) gives for the first two moments

$$\left.\begin{aligned} \mu_1(0) &= \frac{h}{1-\bar{q}}\left[1 - e^{-(1-\bar{q})\,\zeta \operatorname{cosec} b}\right], \\ \mu_2(0) &= 2h^2\left[\frac{1}{(1-\bar{q})(1-\bar{q}^2)} + \frac{1}{(\bar{q}-1)(\bar{q}-\bar{q}^2)}\, e^{-(1-\bar{q})\,\zeta \operatorname{cosec} b} + \right. \\ &\qquad \left. + \frac{1}{(\bar{q}^2-\bar{q})(\bar{q}^2-1)}\, e^{-(1-\bar{q}^2)\,\zeta \operatorname{cosec} b}\right]. \end{aligned}\right\} \tag{86}$$

If all the clouds have the same transparency q, Eqs. (86) take the form

$$\left.\begin{aligned}\mu_1(0) &= \frac{h}{1-q}\left[1-e^{-(1-q)\zeta\operatorname{cosec} b}\right],\\ \mu_2(0) &= \frac{2h^2}{(1-q)^2}\left[\frac{1}{1+q}-\frac{1}{q}e^{-(1-q)\zeta\operatorname{cosec} b}+\frac{1}{q(1+q)}e^{-(1-q^2)\zeta\operatorname{cosec} b}\right].\end{aligned}\right\} \qquad (87)$$

Eqs. (86) take on a particularly simple form for directions in the galactic plane, namely for $b = 0$:

$$\left.\begin{aligned}\mu_1 &= \frac{h}{1-q},\\ \mu_2 &= \frac{2h^2}{(1-q)(1-q^2)}.\end{aligned}\right\} \qquad (88)$$

If we use observational data to calculate $\mu_1(0)$ and $\mu_2(0)$ for various b, Eqs. (87) provide two redundant systems of equations which are solved for q, ζ, and h.

The average absorption in a single dark nebula $\Delta m = -2.5 \lg q$, determined by various authors from brightness fluctuations of the Galaxy and the Metagalaxy (see §187), was found to be $0.20-0^m.30$, and the expectation value ζ of the number of dark clouds in the direction to the galactic poles was found to range between 1 and 1.5.

The relative mean square brightness fluctuation is expressed in terms of the moments by

$$\delta^2 = \frac{\mu_2 - \mu_1^2}{\mu_1^2}.$$

Using Eqs. (87) the reader will now easily prove that σ^2 should increase with decreasing b. This is consistent with the data of Table 13 and shows that the dark clouds are indeed the dominant factor in shaping the brightness fluctuations of the Galaxy.

§131. Main features of galactic structure. To first approximation the Galaxy may be described as a flattened ellipsoid of revolution. This model is justified by the variation of the function $N(m)$ with galactic latitude and longitude, by dynamic considerations and galactic rotation data, and by observations of the shape of other galaxies with stellar populations not unlike our Galaxy.

The plane of symmetry of this ellipsoid, the galactic plane, is inclined at an angle of about 62° to the plane of the celestial equator, and the longitude of the ascending node is $\sim 280°$ (see §111).

The direction to the center of the Galaxy can be found by several methods. 1) Investigation of the density function. The direction in which the star density increases fastest with increasing distance is close to the direction to the galactic center. This follows from the dynamic theory of rotating stellar systems, according to which the star density monotonically decreases with increasing galactocentric cylindrical coordinates R and Z. 2) Determination of the geometrical center of the system of globular clusters. The center of inertia of the system of globular clusters, which has approximately the same size as the Galaxy and is definitely a part of the Galaxy, is apparently close to the center of inertia of the Galaxy. 3) Determination of the kinematic center of rotation of the Galaxy from studies of star motion and motion of other galactic objects. 4) Photographs of the

central parts of the Galaxy taken in 1949 by Kalinyak, Krasovskii, and Nikonov at 9800 Å using an electron-optical image converter and investigation of the brightness distribution at 10,300 Å near the central parts of the the Galaxy carried out by Stebbins and Whitford in 1947. 5) Radio observations at 21.1 cm wavelength of the motion of neutral interstellar hydrogen. 6) Radio observations of the distribution of ionized hydrogen at various wavelengths.

The results of these different methods show a good fit with one another, as we see from Table 14, originally compiled by Oort and subsequently supplemented with new data.

TABLE 14

Method of determination	l_0
Star counts in moderate and high latitudes . . .	324° ± 5°
Globular clusters (Shapley)	325° ± 3
Planetary nebulae (Minkowski)	328° ± 3
Infrared radiation (Stebbins, Whitford)	326° ± 2
Differential rotation (various determinations) .	325 ± 1
Velocities of fast objects	323 ± 4
100 MHz radio waves (Bolton, Westfold)	329 ± 1
200 MHz radio waves (Allen, Gum)	325 ± 1
22 cm and 75 cm radio waves (Westerhout) . . .	327.5 ± (0.5)
21.2 cm radio waves (Kwee, Muller, Westerhout)	327.5 ± (0.5)

The weighted average of these determinations is close to 327° and shows good agreement with the results of infrared and radio observations, which are the most accurate. It should be remembered that the radio emission of the Galaxy gives the center of distribution of interstellar gas. Since the mass of the interstellar gas is only about 0.02 of the entire galactic mass, the center of inertia of the interstellar gas need not coincide with the galactic center of inertia.

According to radio observations at 21.1 cm wavelength, the center of the Galaxy is located at galactic latitude $-1°.5$, radio observations at 22 and 75 cm give latitude of $-1°.1$.

It thus seems that the center of the Galaxy has the coordinates $l = 327°$ and $b = -1°.3$.

The distance to the center of the Galaxy is determined primarily as the distance to the geometrical center of the system of globular clusters. Interstellar absorption must be carefully taken into consideration, as well as the selection effect associated with the fact that some globular clusters possibly lie on the other side of the galactic center and have not yet been discovered. These effects are of no significance when we are concerned with the general direction to the galactic center, but they may markedly distort the distance. Shapley, who first determined the distance of the galactic center from the distribution of globular clusters when the interstellar absorption was not known, obtained a result double the true figure. Selection and incompleteness of the list of globular clusters conversely reduces the distance to the galactic center.

Another method of distance determination is kinematic. The distribution of the proper motions of stars gives the angular velocity ω of the revolution of the Sun around the center of the Galaxy. From the radial velocities of the near galaxies we then find the velocity v of the Sun relative to the system of the near galaxies. If we assume that this velocity is largely due to the motion of the Sun in the Galaxy, and the residual velocity of the Galaxy itself is ignored, the ratio

$$R = \frac{v}{\omega}$$

gives the distance R to the center of the Galaxy in kilometers (v is in km/sec, ω in radians per sec).

Both methods give close results. The distance to the center of the galaxy is approximately 7500 pc.

To determine the major radius of the Galaxy, we require the distance to its rim in the direction of the anticenter, as well as the distance of the galactic center. The galactic rim, however, is not perfectly sharp. The outer boundary can be determined, say, as the locus of points where the stellar density is equal to the star density in the intergalactic space. Since encounters with other stars sometimes impart critical velocities to certain stars, which consequently escape from the galaxies, the stellar density in the intergalactic space is definitely not zero.

The dynamic boundary of the Galaxy is regarded as the curve where a star has a probability of 1 of being detached from the Galaxy during one galactic revolution due to random encounters with other galaxies.

In practice, the boundary of the Galaxy is inferred from the distances of the farthest high-luminosity objects. Thus, some open clusters and planetary nebulae in the region of the galactic anticenter lie at a distance of about 4000 pc. Long-period cepheids in the same direction are found at distances of up to 6000 pc. If the last figure, 6000 pc, is adopted as the distance to the rim of the Galaxy in the direction of the anticenter, the diameter of the Galaxy is $(7500 + 6000) \times 2 = 27{,}000$ pc.

To pinpoint the position of the Sun in the Galaxy, we have to determine its distance from the galactic plane. In §107 we introduced the concept of the galactic equator and tacitly implied that the Sun lies in the plane of symmetry of our stellar system. Otherwise, the plane of concentration of the galactic objects will intersect the celestial sphere over some small, and not great, circle. The galactic equator, on the other hand, is by definition the great circle having a common pole with this small circle. The galactic equator can also be defined as the limiting position of the concentration curve of progressively farther galactic objects.

The position of the Sun relative to the galactic plane is found by determining the average distance $\bar{z}$ of a large number of stars from the plane of the galactic equator. In these calculations, the sign of z, given by $z = r \sin b$, must be taken into consideration. If $\bar{z}$ were equal to zero, this would imply that the Sun lies precisely in the plane of symmetry of the Galaxy. The results generally give $\bar{z}$ between -10 and -20 pc. This shows that the Sun lies north of the galactic plane. It must be remembered, however, that this result may be a reflection of the peculiar structure of the local part of the Galaxy near the Sun. Anyhow, the distance of the Sun from the galactic plane of symmetry is negligibly small compared to the size of the stellar system and is therefore generally ignored.

The surfaces of equal density in the ellipsoidal model of the Galaxy are similar ellipsoids of revolution. The star density near the Sun is 0.120 ± 0.008 stars/pc^3 (see §116). If an average stellar mass is taken to be $0.42\, m_\odot = 8.4 \cdot 10^{32}$ g, this gives $3.5 \cdot 10^{-24}$ g/cm^3. The distribution of material density in the Galaxy is also found from dynamic considerations. According to the analysis (by Kuzmin) of the recent findings of Oort, Kuzmin, Parenago, and Safronov, the density near the Sun is $(5.2 \pm 0.5) \times 10^{-24}$ g/cm^3. The difference between the overall density figure and the density of stellar matter, which is $1.7 \cdot 10^{-24}$ g/cm^3, provides an indication of the density of interstellar matter near the Sun.

According to Kuzmin, the logarithmic density gradient in the plane of the Galaxy near the Sun is $\frac{d \lg \varrho}{dR} = -0.15\,\text{kpc}^{-1}$, and the logarithmic gradient to the galactic pole is $\frac{d \lg \varrho}{dz} = -0.94\,\text{kpc}^{-1}$. The semiaxis ratio of the surfaces of equal density in the Galaxy is thus approximately $10^{-(0.94-0.15)} \approx 0.16$.

§132. Two types of stellar populations. Subsystems of the Galaxy. The model of the Galaxy consisting of surfaces of equal density in the form of similar ellipsoids of revolution is an adequate approximation for the treatment of dynamic problems. From the viewpoint of stellar statistics, however, this model is much too crude, since it ignores the discrete nature of the stars. An exact galactic model should assign to each point not only the star density but also an appropriate distribution functions of stellar luminosities, spectra, and other characteristics at that point.

In 1944, proceeding from a study of the distribution of various objects in our stellar system and in other galaxies, W. Baade established the existence of two types of stellar population. Type I population includes the hot giants and supergiants, long-period cepheids, novae and supernovae, open clusters, hydrogen clouds, and dust nebulae. Stars of Population I are located near the symmetry plane of their spiral galaxy, concentrating in the spiral arems and avoiding the nucleus. Population II stars are red dwarfs, red giants, short-period cepheids, and globular clusters. The globular clusters are exclusively composed of Population II stars. Stars of this population form the galactic nuclei; in spiral galaxies they are dominant in regions far from the galactic plane; in elliptic galaxies they are the only stellar component.

Objects of Populations I and II as if avoid one another: long-period cepheids, for instance, do not occur in globular clusters, where numerous short-period cepheids are observed. Globular clusters and elliptic galaxies are rich in red giants, but have no hot supergiants, dust and gas. The red dwarfs are apparently an exception to this avoidance rule: they are omnipresent, occurring in all parts of the Galaxy.

An interesting distinctive characteristic of the two populations is the luminosity per unit mass: for Population I stars L/m is on the average of the order of 100 erg/sec/g, being much higher than for Population II stars, where this parameter is on the average of the order of 0.1 erg/sec/g, L/m can thus be accepted as a quantitative identifying characteristic of the stellar population in a particular part of the star system.

In 1942, Kukarkin, Parenago and other astronomers of the Moscow school suggested that the Galaxy can be regarded as an assembly of subsystems comprising objects with definite physical characteristics. For

example, if we ignore all the galactic objects except the Mira Ceti variables, we are left with the subsystem of Mira Ceti stars. Different subsystems are intertwined and interpenetrating. Each subsystem is morphologically close to an ellipsoid of revolution. It is significant that different subsystems show different flattening. They are accordingly divided into three groups. The first group, characterized by the maximum flattening, forms the disk component of the Galaxy. The second group comprising less flattened subsystems is the intermediate component. The third group of almost spherical subsystems is the so-called halo component of the Galaxy. The objects of the disk component show stronger galactic concentration than the objects of the intermediate component, whereas the halo component shows almost no galactic concentration. On the other hand, the halo component shows the maximum concentration toward the center of the Galaxy, whereas the disk component actually avoids the center. This fact is in good agreement with the theoretical results of the dynamics of rotating self-gravitating objects.

The first subsystem characteristics were obtained by Kukarkin for the variable stars. The variables have certain advantages compared to other stellar objects: 1) these are mostly high-luminosity objects visible over large distances; 2) they are readily distinguishable among the stationary stars; 3) they are easily classified according to their light variation curve.

The variables were found to include representatives of all the three galactic components. Their characteristics were investigated by Parenago, Ikaunieks, and others.

Quantitatively, each subsystem can be characterized by three parameters: 1) the total number of objects in the subsystem, 2) the logarithmic density gradient of the objects in the galactic plane $\frac{d \lg D}{dR}$, which characterizes the concentration toward the center of the Galaxy, and 3) the logarithmic density gradient at right angles to the galactic plane, $\frac{d \lg D}{dz}$, which characterizes the concentration toward the plane of the Galaxy. The logarithmic gradients are constant at any point if the density varies according to the barometric law. In this case (see §120), the half-thickness β of the equivalent homogeneous layer of the subsystem is $\frac{\lg e}{\frac{d \lg D}{dz}}$.

It follows from dynamic considerations that the flattening of each subsystem should be related to the mean velocity component perpendicular to the galactic plane. In disk subsystems this velocity component should be relatively small, since otherwise the objects would escape to large distances from the galactic plane and the system would lose its planar disk shape. Determination of the dispersion in velocities σ_z at right angles to the plane of the Galaxy confirms the expected dependence of σ_z on the flattening of the subsystem.

Some characteristics of the galactic subsystems are listed in Table 15, which is borrowed from P. P. Parenago's "Course in Stellar Astronomy" (Gostekhizdat. 1954).

The model of the Galaxy comprising interpenetrating subsystems has an obvious advantage over the model of two populations: it ensures a more

detailed study and gives definite quantitative results. An obvious shortcoming is that it ignores the peculiar structural features of the Galaxy, the presence of a nucleus and spiral arms.

TABLE 15

Subsystems	$-\frac{\partial \lg D}{\partial R}$	$-\frac{\partial \lg D}{\partial z}$	β	Number of known objects	Assumed total number	Source
Disk component						
Long-period cepheids	0.11	9.86	44	500	30,000	Kukarkin, 1947
B stars	?	9.5	46	10,000	150,000	Shmirel'man, 1952
Open clusters	0.11	8.2	53	500	33,000	Barkhatova, 1949
O stars	0.10	7.5	58	200	6,500	Chernova, 1948
Dark (dust) nebulae	(0.16)	4.34	100	10^4	10^8	Parenago, 1945
Supernovae	0.15:	5.1:	85:	5	—	Kopylov, 1954
Intermediate component						
Novae	0.22	2.39	182	100	10^6	Kopylov, 1954
RV Tau variables	0.23	2.2	200	80	10^4—10^5	Pavlovskaya, 1950
R and N stars, irregular variables	(0)	4.3	100	130	13,000	Ikaunieks, 1952
R and N stars, semiregular variables	(0)	2.5	175	45	4,500	Ikaunieks, 1952
Constant-magnitude R and N stars	(0)	1.3	335	260	16,000	Ikaunieks, 1952
Planetary nebulae	0.4:	2.0	217	350	130,000	Vorontsov-Vel'yaminov, 1950
S stars	0.14	1.20	360	90	5,000	Ikaunieks, 1952
White dwarfs	0.36:	(1.08)	(400)	100	$5 \cdot 10^9$	Parenago, 1949
Long-period variables	0.26:	0.86	500	3000	$1.3 \cdot 10^6$	Kukarkin, 1947
Halo component						
Long-period variables with $P=150-200$ days	(0.26)	0.22	2000	300	10^5	Kukarkin, 1947
Short-period cepheids	0.27	0.22	2000	3000	170,000	Kukarkin, 1947
Long-period W Vir cepheids	0.25	0.2:	2000:	50	20,000	Kukarkin, 1947
Subdwarfs	0.27	(0.17)	(2600)	300	10^{11}	Parenago, 1949
Globular clusters	0.26	0.15	3000	100	250	Parenago, Kukarkin, Florya, 1941
All stars on the average	(0.25)	1.15	380	10^9	$1.2 \cdot 10^{11}$	Parenago, 1947

§133. The spiral structure of the Galaxy. The conviction that our Galaxy has a spiral sturucture is largely based on comparison with other galaxies of similar stellar composition and comparable size. These are the Sb and Sc galaxies with prominent spiral structure. Another indication of spiral arms in the Galaxy is provided by the substantial inhomogeneity of its structure, associated with the peculiar distribution of interstellar gas, large clouds of dark matter, and the hot high-luminosity stars. These are the objects found in the spiral arms of other galaxies.

One of the main problems of stellar astronomy is to arrive at a reliable charting of the spiral arms in our Galaxy and to fix the position of the Sun relative to these arms. This is a difficult undertaking, since we are viewing the spiral structure from the inside, so that it is not clearly

visible. To determine the true structure of the Galaxy, we should pick out the characteristic objects of spiral arms, find their spatial distribution in the Galaxy, and thus establish the configuration of the spiral arms. The currently available results are uncertain owing to the inaccurate determination of the distances, the screening effect of the abundant dark matter, and the high sensitivity of the results to observation errors.

Because of the numerous errors, the observed distribution of the typical objects gives several alternative arm configurations. The uncertainty becomes especially clear if we remember that the spiral arms may have several turns each. Different studies of the spiral structure therefore led to outright contradictory results, and the true spiral configuration is not clear to this day.

The spiral structure was charted from the distribution of O and B stars (Vorontsov-Vel'yaminov), stellar associations (Morgan, Whitford, Code), emission hydrogen clouds (Gaze, and also Morgan, Sharpless, and Osterbrock), emission hydrogen clouds and open clusters (Weaver), radio hydrogen fields (Shklovskii), radio hydrogen fields at 21 cm wavelength (Oort, van de Hulst, Muller, and also Parenago).

The great advantage of the radio methods is that they are free from absorption errors: the interstellar medium is virtually transparent at radio frequencies.

The existence of the galactic nucleus, like the spiral arms, has been mainly inferred from observations of other galaxies. All spiral galaxies have a prominent nucleus characterized by increased stellar density and monotonic increase of brightness toward the center and devoid of any individual features.

In 1949 Kalinyak, Krasovskii, and Nikonov photographed the galactic nucleus in infrared light at 9800 Å. The angular diameter of the nucleus was $10-12°$, which at a distance of 7500 pc corresponds to a diameter of $1300-1600$ pc. This is the typical size of the nucleus in Sb spiral galaxies.

§134. Radio structure of the Galaxy. Highly valuable information on galactic structure has been recently obtained from radio observations. Strictly speaking, radio observations only reveal the distribution and motion of hydrogen in the Galaxy. However, since the distribution of hydrogen is correlated with the distribution of stars and interstellar dust, it provides an indication of the overall structure of the Galaxy. It is fairly certain, for instance, that the plane and the axis of symmetry of the interstellar hydrogen subsystem coincide with the plane and the axis of symmetry of the other galactic subsystems.

The distribution of neutral hydrogen (so-called HI regions) is observed at 21 cm wavelength, at which the neutral hydrogen has a sharp and unique emission line. Neutral hydrogen also shows strong absorption at this wavelength, which however does not prevent receiving the 21 cm emission from distant HI regions: on account of the differential galactic rotation (see §149), the hydrogen clouds which lie at different distances along the same line of sight move with different radial velocities. The Doppler effect therefore produces a relative shift of the 21 cm line of the different hydrogen clouds, so that the near HI regions become transparent to the shifted emission of the distant hydrogen clouds. The Doppler shift, combined

with galactic rotation data (see §151), makes it possible to determine the distance of the radio source, and the radiation intensity at 21 cm wavelength gives the density of neutral hydrogen at the given point.

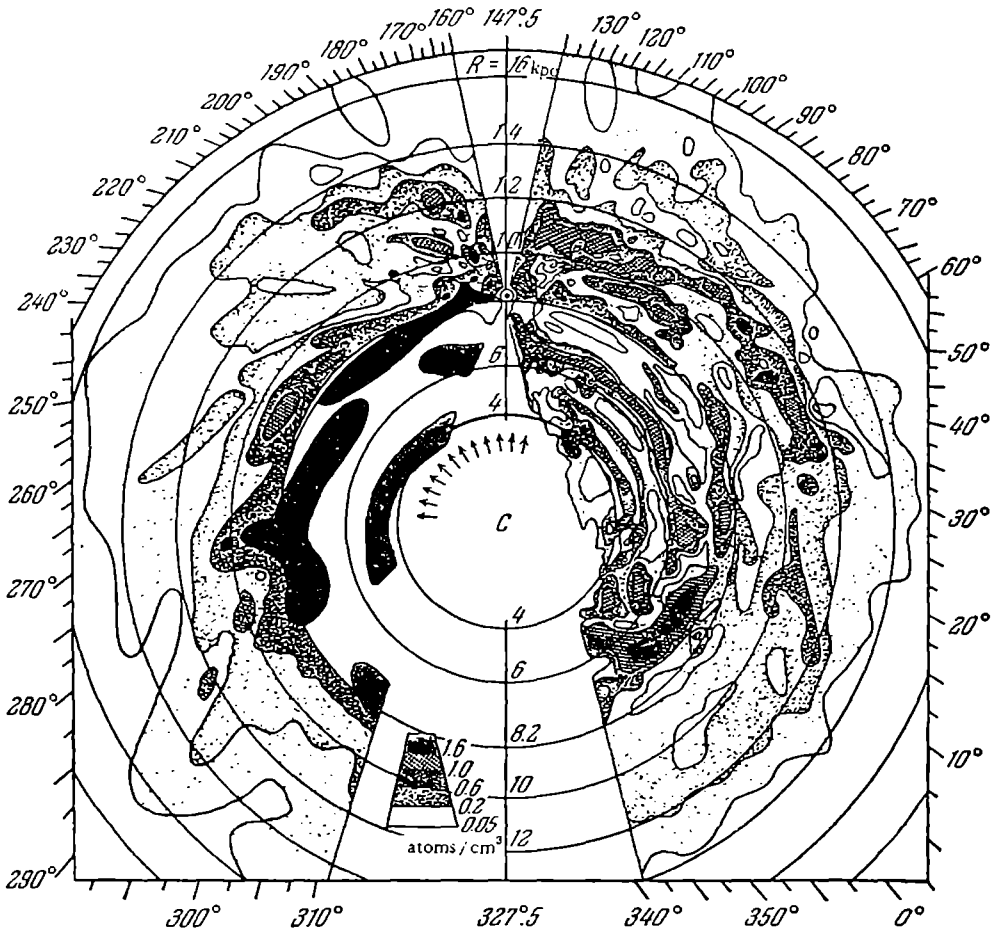


FIGURE 163

A detailed study of the distribution of neutral hydrogen in the Galaxy was carried out in Leiden by van de Hulst, Muller, and Oort in 1954 and in Sydney by Kerr, Hindman and Stahr-Carpenter in 1957. These surveys covered almost the entire sky and led to the following conclusions, published in 1958 by Oort, Kerr, and Westerhout: 1) The neutral hydrogen is distributed in a thin sheet in the Galaxy. Its scale height (i.e., the thickness that the sheet of the same mass would have if its density at any point of the cross section were equal to the maximum density attainable at the center of the real layer) is approximately 220 pc. 2) The neutral hydrogen sheet in the main closely follows the galactic plane. Within a radius of 10 kpc from the center of the Galaxy, points of maximum density in

cylindrical columns perpendicular to the galactic plane are nowhere distant more than 75 pc from the plane of the Galaxy. Within a radius of 6 kpc, these deviations are even smaller and do not exceed 30 pc. In Oort's opinion, this property of the neutral hydrogen gives a more accurate determination of the position of the galactic pole than star counts do. At the periphery, the neutral hydrogen layer is fairly distant from the galactic plane. Outside the 12 kpc radius, deviations of up to 600 pc from the plane of the Galaxy were observed. In these outer regions, the density of neutral hydrogen is much less than in the interior part of the Galaxy. 3) The parameters characterizing the galactic rotation of neutral hydrogen show a good fit with the rotation parameters of the disk subsystems. 4) The observed density distribution of neutral hydrogen near the plane of the Galaxy (Figure 163) is highly sensitive to the adopted value of the galactic rotation parameters near the Sun (the position of the Sun in the figure is marked by a circle at a distance of 8.2 kpc from the center of the Galaxy in the direction of the longitude 147°.5). Since the relative errors in these parameters are substantial, the hydrogen distribution chart is not particularly reliable. Moreover, it is distorted by residual velocities of hydrogen clouds, which may be substantial, but cannot be accounted for with any accuracy. Because of these distortions, we cannot at this stage fix the position of the spiral arms from neutral hydrogen radio data. There is furthermore a hitherto uncharted region between galactic longitudes of 310 and 340°. These uncertainties notwithstanding, Figure 163 shows in clear outline elongated arches of increased hydrogen density, which are apparently parts of spiral arms. 5) The density of neutral hydrogen is the highest at a distance of 6.5 kpc from the center of the Galaxy, where it reaches 1 atom per 1 cm^3. At a distance of 14 kpc the density falls to 0.1 atom per 1 cm^3. The logarithmic density gradient of neutral hydrogen in the plane of the Galaxy is thus less than the logarithmic density gradient of stars ($-0.13\ kpc^{-1}$ and $-0.15\ kpc^{-1}$, respectively). The density of neutral hydrogen in the interior of the Galaxy is less than the density of stellar matter, but as it falls off at a slower rate toward the periphery, the density of neutral hydrogen becomes comparable to the density of stellar matter in the outer regions. 6) The neutral hydrogen within a sphere of 3 kpc radius from the center of the Galaxy reveals substantial outward velocities, reaching 130 km/sec. This creates an impression of steady efflux of hydrogen from the center of the Galaxy to the periphery.

TABLE 16

R, kpc	$\bar{n}_H$, atoms in 1 cm^3	R, kpc	$\bar{n}_H$, atoms in 1 cm^3
3—4	0.33	9—10	0.41
4—5	0.72	10—11	0.55
5—6	0.68	11—12	0.40
6—7	0.97	12—13	0.21
7—8	0.72	13—14	0.09
8—9	0.48	14—15	0.04

Muller and Westerhout in 1957 compiled a catalogue of the 21 cm line profiles for neutral hydrogen at 694 points within the area $-10° < b < +10°$, $220° < l < 318°$. The profiles give the emission intensity as a function of radial velocity. Using this catalogue, Westerhout determined the average density $\bar{n}_H$ of neutral hydrogen as a function of distance from the center in the galactic sector with the Sun at its apex (Table 16). The density maxima at 6—7 kpc and 10—11 kpc are possibly an indication of spiral arms.

Ionized hydrogen (so-called HII regions) emits a continuum of radio frequencies. It also absorbs the corresponding wavelengths. The kinetic

temperature of ionized hydrogen is estimated at 10,000°K. Ionized hydrogen is therefore observed in absorption if the background emission temperature is over 10,000°K and in emission if the background emission temperature is less than 10,000°K. We know from observations that at wavelengths over 10 m, the background temperature is greater than 10,000°K; therefore at these wavelengths the HII regions are observed in absorption. Shain, using the Sydney radio telescope in 1957, established the existence of large HII regions both near the plane of the Galaxy and at high galactic latitudes. Many of these radio clouds are identified with optically observed HII regions.

At wavelengths shorter than 10 m, the HII regions are observed in emission. The intensity distribution at these wavelengths was studied by Mills, Hill, and Slee in 1958 (meter wavelengths) and by Westerhout in 1958 (75 cm and 22 cm). The observational difficulties at these wavelengths are associated with the existence of an alternative radio source of nonthermal origin at meter wavelengths. According to Shklovskii, this nonthermal emission is the radiation of relativistic electrons in a magnetic field.

Westerhout found that the maximum radiation intensity at 22 cm is observed in the plane of the Galaxy at galactic longitude 353°, decreasing toward the center. He attributed this fact to the presence of an ionized hydrogen ring with a radius of 3—4 kpc around the center of the Galaxy. Westerhout's observations extended only to galactic longitude 320°, and therefore the other part of the ring was not observed.

The ionized hydrogen radiation is proportional both to the density of ions and the density of electrons, so that it is proportional to the square of the density of interstellar matter. This shows that the ionized hydrogen collected into isolated clouds emits more than the continuous distribution of interstellar hydrogen within the same volume. The radiation intensity increases with increasing cloud density. Assuming that the HII regions have an average density of 5 atoms/cm^3 and the clouds near the Sun occupy 0.0015 of the total space volume, Westerhout obtained the following ratios of the mean ionized hydrogen density to the mean neutral hydrogen density in the plane of the Galaxy for various distances from the galactic center (Table 17).

TABLE 17

R, kpc	n_H^+/n_H	R, kpc	n_H^+/n_H
0—2	0	5—6	0.09
2—3	0.09	6—7	0.06
3—3.5	0.36	7—8	0.04
3.5—4	0.64	8—9	0.06
4—4.5	0.21	9—10	0.07
4.5—5	0.13	>10	0

The total mass of neutral hydrogen is close to $1.4 \cdot 10^9$ solar masses and is about 0.02 of the galactic mass. The mass of ionized hydrogen is $8 \cdot 10^7$ solar masses, which is about 0.0001 of the galactic mass.

Westerhout also carried out measurements at 22 and 75 cm near the galactic center and charted the surface brightness isophotes. At both wavelengths, a galactic nucleus 4—5° in diameter is prominent, centered at $l = 327°.5$, $b = -1°.1$.

A study of the distribution of nonthermal radio emission led Shklovskii to the conclusion that the Galaxy is surrounded by a spherical corona of highly tenuous hydrogen. Pikel'ner reached the same conclusion from different considerations.

§135. The Local Group. In §130 the galactic equator was defined as the limiting position of the concentration curve of galactic objects at ever increasing distances.

If we consider the nearest stars, the plane of their maximum concentration does not coincide with the plane of the galactic equator. Back in 1879, Gould found that the plane of maximum concentration of stars up to 4^m makes an angle of 17° with the galactic plane. As we proceed to progressively fainter stars, this inclination angle decreases, dropping according to Seares to 8°.1 for stars to 9^m and to 2°.7 for stars to 18^m. The longitude of the ascending node of the plane of concentration thus also changes. For stars up to 4^m it is 257°, for stars up to 9^m it is 5°, and for stars up to 18^m it is 80°.

A number of authors regard these changes as evidence of the existence of a certain physical star system — a stellar cloud — in the part of the Galaxy around the Sun. This so-called Local Group is flattened, but its principal plane does not coincide with the plane of the Galaxy. Examining bright stars only, we focus our attention mainly on the nearest stars of the Local Group, and the inclination of the plane of maximum concentration is therefore almost equal to the inclination of the plane of the Local Group to the galactic plane. As we progress to fainter and therefore more distant stars, stars outside the Local Group enter the sample and the inclination angle of the concentration plane decreases, approaching zero.

Studies of star density near the Sun reveals a marked increase in density in directions from $l = 230°$ to $l = 260°$ up to distances of 300 — 400 pc. This distance is accepted as the approximate radius of the Local Group. It is significant that the increase in density applies mainly to B-type and partly to A-type stars. We are thus dealing here not with increased overall density, but with increased partial density of B and partly A stars. If this is so, the Local Group is physically unstable due to insufficient overall star density. On the other hand, the rotation data obtained by Ogorodnikov and Shatsova seem to indicate that the Local Group is stable. According to their findings, the rotation speed of the Local Group is a factor of 2 — 3 greater than the angular rotation velocity of the Galaxy.

REFERENCES TO CHAPTER XVIII

1. AGEKYAN, T.A. Determination of the density function and the absorption coefficient in the plane of the Galaxy.— Uchenye Zapiski, Leningrad University, 18, No. 116. 1949.
2. AGEKYAN, T.A. Estimating the total number of stars in the Galaxy by extrapolation of the magnitude function.— Uchenye Zapiski, Leningrad University, 18, No. 116. 1949.
3. AGEKYAN, T.A. Theory of brightness fluctuations of the Milky Way and the Metagalaxy.— Vestnik, Leningrad University, No. 1. 1956.
4. AMBARTSUMYAN, V.A. The relationship between diffuse nebulae and exciting stars.—Doklady AN Armenian SSR, 2, No. 3. 1945.
5. AMBARTSUMYAN, V.A. Relative distribution of luminous and dark matter in the Galaxy.— Doklady AN Armenian SSR, 4, No. 5. 1946.
6. AMBARTSUMYAN, V.A. Surface brightness in the Galaxy.— Astron. Zh., 23, No. 5. 1946.
7. AMBARTSUMYAN, V.A. Theory of fluctuations in the apparent distribution of stars in the sky.— Bull. Byurakan Obs., 4. 1951.

8. BAADE, W. Resolution of Messier 32, NGC 205, and the central region of the Andromeda Nebula, NGC 147 and NGC 185, two new members of the Local Group of galaxies.— Astron. Zh., 22, No. 3. 1945 [translation into Russian].
9. BOK, B. and P. BOK. The Milky Way.— Harvard Univ. Press. 1957.
10. VASHAKIDZE, M.A. A study of the galactic absorption from color excess of extragalactic nebulae and long-period cepheids and by other methods.— Bull. Abastumani Astrophys. Obs., No. 13, 1953.
11. VORONTSOV-VEL'YAMINOV, B.A. Clouds of hot giants and Milky Way clouds.— Astron. Zh., 30, No. 4. 1953.
12. VORONTSOV-VEL'YAMINOV, B.A. The spiral structure of the Galaxy.— Astron. Zh., 30, No. 1. 1953.
13. SONN, W. and K. RUDNITSKI. Stellar Astronomy.— [Russian translation, 1959.]
14. IKAUNIEKS, Ya. Ya. Spatial distribution and kinematics of carbon stars.— Astron. Zh., 29, No. 6. 1952.
15. IKAUNIEKS, Ya. Ya. Spatial distribution of irregular stars.— Peremennye Zvezdy, 8, No. 6. 1952.
16. KALINYAK, A.A., V.I. KRASOVSKII, and V.B. NIKONOV. Observations of the galactic center in infrared light.— Doklady AN SSSR, 66, No. 1. 1949.
17. KUZMIN, G.G. Mass distribution in the Galaxy.— Publ. Tartu Astron. Obs., 32, No. 4. 1952.
18. KUZMIN, G.G. The value of the dynamic parameter c and the density near the Sun.— Publ. Tartu Astron. Obs., 33, No. 1. 1955.
19. KUKARKIN, B.V. Issledovanie stroeniya i razvitiya zvezdnykh sistem na osnove izucheniya peremennykh zvezd (Investigation of the structure and evolution of stellar systems from a study of variable stars.— Gostekhizdat. 1949.
20. KUNITSKII, R.V. The system of the nearest B and A stars.— Astron. Zh., 13, No. 1. 1943.
21. MARKARYAN, B.E. Fluctuations in the apparent distribution of stars and interstellar absorption.— Bull. Byurakan Obs., No. 1. 1946.
22. PARENAGO, P.P. Studies based on the Shternberg Institute catalogue of stellar parallaxes.— Trudy, Shternberg Astronomical Institute, 13, No. 1. 1940.
23. PARENAGO, P.P. Dark nebulae and absorption of light in the Galaxy.— Astron. Zh., 17, No. 4. 1940; 22, No. 3. 1945.
24. PARENAGO, P.P. Kurs zvezdnoi astronomii (A Course in Stellar Astronomy).— Gostekhizdat. 1954.
25. PARENAGO, P.P. A study of the mass — luminosity relation.—Astron. Zh., 14, No. 1. 1937.
26. RUSAKOV, G.I. Brightness fluctuations of the Milky Way and physical characteristics of diffuse nebulae.— Uchenye Zapiski, Leningrad University (phys. math. series), 18, No. 116. 1949.
27. SAFRONOV, V.S. Density of matter in the Galaxy near the Sun.— Astron. Zh., 29, No. 2. 1952.
28. FESENKOV, V.G. The number of stars in the galactic system.— Doklady AN SSSR, new series, 28, No. 7. 1940.
29. FIORYA, N.F. A study of interstellar absorption.— Trudy, Shternberg Institute, 16. 1946.
30. KHABIBULLIN, Sh.G. Distribution of star densities in high galactic latitudes.— Astron. Zh., 26, No. 5. 1949.
31. KHARADZE, E.K. Color index catalogue of 14,000 stars and the study of interstellar absorption in the Galaxy from color excesses.— Bull. Abastumani Astrophys. Obs., No. 12. 1952.
32. KHOLOPOV, P.N. On the nucleus of the Galaxy.— Astron. Zh., 27, No. 2. 1950.
33. SHATSOVA, R.B. Luminosity function of absolutely weak stars.— Astron. Zh., 29, No. 5. 1952.
34. AMBARZUMIAN, V.A. and Sh. G. GORDELADSE. Problem of diffuse nebulae and cosmic absorption.— Bull. Abastumani Astrophys. Obs., No. 2. 1938.
35. BECKER, W. Sterne und Sternsysteme. 1942.
36. BOK, B.J. The distribution of the stars in space.— Chicago. 1937.
37. BRILL, A. Über die Losungen der Integralgleichung der Stellarstatistik.— Astron. Nachr., 257. 1935.
38. CEDERBLAD, S. Studies of bright diffuse galactic nebulae with special regard to their spatial distribution.— Lund Medd., 12, No. 119. 1946.
39. HUBBLE, E. The general study of the diffuse galactic nebulae.—Mt. Wilson Contrib., No. 241. 1922.
40. VAN de HULST, C., C.A. MULLER, and J.H. OORT. The spiral structure of the outer part of the galactic system derived from the hydrogen emission at 21 cm wavelength.— Bull. of the Astr. Inst. of the Netherlands, 12, 117 (No. 52). 1954.
41. KUIPER, G.P. The nearest stars.— ApJ, 95, 201 – 212. 1942.
42. OGORODNIKOFF, K. Theoretical analyses of stars counts in obscured regions.— Izvestiya Pulkovo Obs., 16, No. 3 (123). 1949.

43. OORT, J.H. Absorption and density distribution to the galactic system.— Bull. Astron. Inst. Netherl., 8, No.308. 1938.
44. OORT, J.H. Some problems concerning the structure and dynamics of the galactic system and the elliptical nebulae NGC 3115 and 4494.— ApJ, 94, 318. 1940.
45. OORT, J.H. Problems of galactic structure.— ApJ, 116, No.2. 1952.
46. OORT, J.H., G.I. KERR, and G. WESTERHOUT. The galactic system as a spiral nebula.— MN, 118, 379. 1958.
47. PAHLEN, E. Lehrbuch del stellarastronomie.— Leipzig, 1937.
48. VAN RHIJN, P. Distribution of stars according to apparent magnitude, galactic latitude and galactic longitude.— Groningen Publications, No.43, 1929.
49. SEARES, F.H. A numerical method of determining the space density of stars.— ApJ, 74, 268—287. 1931.
50. SEARES, F.H., P. VAN RHIJN, et al. Mean distribution of stars according to apparent magnitude and galactic latitude.— ApJ, 62, 320—374. 1925.
51. SEARES, F.H. and M.C. JOYNER. Systematic deviation from the mean stellar distribution.—ApJ, 67, 24—82. 1928.
52. SMART, W.M. Stellar Dynamics.— Cambridge Univ. Press. 1938.
53. TRUMPLER, R.G. and H.F. WEAVER. Statistical Astronomy.— Univ. of California Press. 1953.
54. VASHAKIDSE, M.A. The space distribution of F type stars.— Bull. Abastumani Astrophys. Obs., No.1. 1937.
55. WEAVER, H.F. B-stars and spiral structure in the larger neighborhood of the Sun.— Astron. J., 58, No.1211. 1953.

A.N. DEICH and T.A. AGEKYAN

Chapter XIX

STELLAR MOTIONS

§136. General considerations. The motion of a star in space can be determined in relation to other stars or stellar systems. It is naturally different relative to different star groups. Inside the Galaxy, the most general description of star motion is that relative to all the stars of the galactic system. In this case let $\dot{X}_*, \dot{Y}_*, \dot{Z}_*$ be the projections of the spatial velocity of a star on the three rectangular axes. For the center of inertia of all the stars in the Galaxy we have by definition

$$\sum \dot{X}_* = \sum \dot{Y}_* = \sum \dot{Z}_* = 0,$$

where the sum is taken over all the stars of our stellar system.

The motion of a star relative to the center of inertia cannot be found, since the motions of the stars in the Galaxy are not all known. It could be determined if we were to measure the motions of a group of objects distributed symmetrically about the galactic center. Assuming symmetric motion of stars relative to the center of the Galaxy, we can regard this point as the center of inertia of the system and refer all the spatial velocities of stars to the galactic center.

Since the observations are carried out from the Earth, we measure the difference in the motions of the Sun and the star (after correction for the Earth's orbital motion). The projections of this difference on the three rectangular axes are designated $\dot{X}, \dot{Y}, \dot{Z}$. Then

$$\left.\begin{aligned} \dot{X}_* - \dot{X}_\odot &= \dot{X}, \\ \dot{Y}_* - \dot{Y}_\odot &= \dot{Y}, \\ \dot{Z}_* - \dot{Z}_\odot &= \dot{Z}. \end{aligned}\right\} \qquad (1)$$

Adding up these equalities for all the stars and seeing that for the center of inertia $\sum \dot{X}_* = \sum \dot{Y}_* = \sum \dot{Z}_* = 0$, we obtain for the projections of the Sun's motion

$$\left.\begin{aligned} \dot{X}_\odot &= -\frac{1}{N} \sum \dot{X}, \\ \dot{Y}_\odot &= -\frac{1}{N} \sum \dot{Y}, \\ \dot{Z}_\odot &= -\frac{1}{N} \sum \dot{Z}, \end{aligned}\right\} \qquad (2)$$

where N is the number of stars in the Galaxy. Having thus determined the velocity of the Sun, we use Eqs. (1) to find the velocities of all the stars relative to the center of the Galaxy.

In practice, we can observe only stars near the Sun. The velocities of the Sun and the stars are then referred to the local center of inertia of stellar motions. Stars near the Sun, as we will see from the following, largely share the motion of the Sun relative to the galactic center. Therefore this common velocity cancels out when we take the difference in the motions of the Sun and the stars and the residual velocity is found to be relatively small. The velocity of the Sun relative to the local center of inertia is approximately 1/10 of its velocity relative to the galactic center.

Let the projections of the velocity of the Sun and stars relative to the local center of inertia be $\dot{x}, \dot{y}, \dot{z}$. Taking $\Sigma\dot{x}_* = \Sigma\dot{y}_* = \Sigma\dot{z}_* = 0$ we find as before

$$\left.\begin{aligned} \xi_\odot &= -\frac{1}{n}\sum \dot{x}_* \\ \eta_\odot &= -\frac{1}{n}\sum \dot{y}_* \\ \zeta_\odot &= -\frac{1}{n}\sum \dot{z}_* \end{aligned}\right\} \qquad (3)$$

where n is the number of observed stars. For each star we have

$$\left.\begin{aligned} \dot{x}_* &= \dot{x} - \dot{x}_\odot \\ \dot{y}_* &= \dot{y} - \dot{y}_\odot \\ \dot{z}_* &= \dot{z} - \dot{z}_\odot \end{aligned}\right\} \qquad (4)$$

The spatial velocity of stars is expressed in km/sec.

§137. **The observed spatial velocity components and their relation to $\dot{x}, \dot{y}, \dot{z}$.** Direct observations do not give the spatial velocities of stars. The observations give the projection of this velocity on the line of sight, called the radial velocity v_r (km/sec), and the proper motion components on the celestial sphere μ_α and μ_δ, which are expressed in seconds of arc per year and are directed along the right ascension and declination of the star. μ_α is often replaced by $\mu \cos\delta$, which is directed along the great circle through the star at right angles to its circle of declination. In photographic measurements the components $\mu_\alpha \cos\delta$ and μ_δ are denoted by μ_x and μ_y and are expressed in seconds of arc. The total proper motion μ is $\sqrt{\mu_x^2 + \mu_y^2}$, and its position angle ψ is obtained from the relation $\operatorname{tg}\psi = \frac{\mu_x}{\mu_y}$. We may also write $\mu = \mu_x \sin\psi + \mu_y \cos\psi$. The position direction of μ_x is taken in the direction of increasing right ascension, i.e., from west to east, and μ_y is reckoned in the direction of increasing declination, i.e., from south to north. The radial velocity is taken positive if the star moves away from the observer.

To determine the spatial velocity, we have to change over from the proper motion μ to the tangential velocity v_t, i.e., the velocity component at right angles to the line of sight, also expressed in km/sec. This requires knowledge of the distance r to the star (in parsecs).

If the distance is measured in parsecs, the tangential velocity is $v_t = k\mu r$, where

$$k = 4.74 = \frac{149\,600\,000\ \text{km}}{31\,556\,926\ \text{sec}}.$$

Here the numerator is the radius of the Earth's orbit in kilometers and the denominator the number of seconds in a tropical year. Indeed, we see from Figure 164 that $v_t = r\mu'' \sin 1''$ pc annually. One parsec is equal to $\frac{149\,600\,000 \text{ km}}{\sin 1''}$, and to obtain the tangential velocity in km per second of time this number should be divided by the number of seconds in a year.

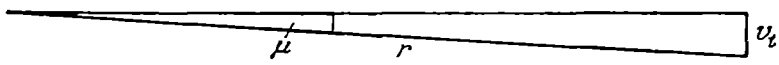


FIGURE 164

The spatial velocity of a star is $v = \sqrt{v_r^2 + v_t^2}$.

Thus, in order to determine the spatial velocity of a star we require v_r, μ, and r, all of which are available from observations. The measurement of stellar distances r or, alternatively, stellar parallaxes $p = \frac{1}{r}$ is the least reliable and therefore introduces the largest error in v. Let the x, y, z system be oriented with the axis x directed to the point of vernal equinox, the axis y lying in the equatorial plane turned through 90° in the direction of increasing right ascension, and the axis z pointing to the north pole. The direction cosines between the systems μ_λ, μ_β, v_r and x, y, z are functions of the equatorial coordinates α and δ of the star:

	$kr\mu_\lambda$	$kr\mu_\alpha$	v_r
x	$\sin\alpha$	$\cos\alpha\sin\delta$	$\cos\alpha\cos\delta$
y	$\cos\alpha$	$\sin\alpha\sin\delta$	$\sin\alpha\cos\delta$
z	0	$\cos\delta$	$\sin\delta$

These relations are readily derived from Figure 165. We have

$$\left.\begin{aligned} \dot{x} &= -kr\mu_x \sin\alpha - kr\mu_y \cos\alpha \sin\delta + v_r \cos\alpha\cos\delta, \\ \dot{y} &= kr\mu_\lambda \cos\alpha - kr\mu_\beta \sin\alpha\sin\delta + v_r \sin\alpha\cos\delta, \\ \dot{z} &= kr\mu_\beta \cos\delta + v_r \sin\delta. \end{aligned}\right\} \tag{5}$$

and conversely

$$\left.\begin{aligned} \mu_\lambda &= \mu_\alpha \cos\delta = \frac{-\dot{x}\sin\alpha + \dot{y}\cos\alpha}{kr}, \\ \mu_\beta &= \mu\delta = \frac{-\dot{x}\cos\alpha\sin\delta - \dot{y}\sin\alpha\sin\delta + \dot{z}\cos\delta}{kr}, \\ v_r &= \dot{x}\cos\alpha\cos\delta + \dot{y}\sin\alpha\cos\delta + \dot{z}\sin\delta, \end{aligned}\right\} \tag{6}$$

$$k = 4.738$$

The spatial velocity components μ_λ, μ_β, and v_r are distorted by observation and measurement errors which are considered in detail in Vol. I of the Course (see also /1/). The observed radial velocity of a star, corrected for the Earth's motion around the Sun and around its axis, is in fact

the line-of-sight projection of the velocity components $\dot{x}, \dot{y}, \dot{z}$ relative to the Sun. These velocities incorporate the effect of galactic rotation and for some stars the K-effect as well. This problem is discussed in §144 and §149.

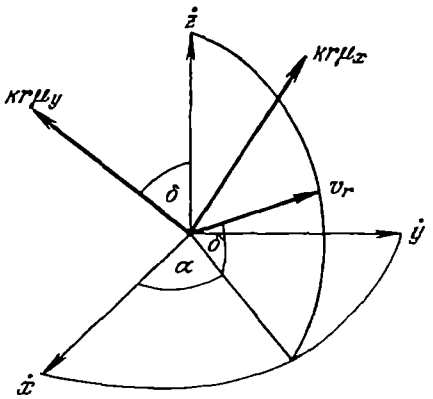


FIGURE 165

The proper motion of a star in the sky includes not only the so-called peculiar motion intrinsic of the star itself but also its parallactic displacement, which is a reflection of the Sun's motion. The proper motion of a star is thus its motion relative to the Sun, which again fits the definition of the components $\dot{x}, \dot{y}, \dot{z}$. The galactic rotation effect is incorporated in the proper motion.

The determination of proper motions meets with a certain difficulty which is not observed in the determination of radial velocities. The point is that the angle which is a measure of the proper motion is the angle between two directions to the star observed at different times. We therefore require some fixed directions in space from which the direction to the star can be taken. In fundamental astrometry these "fixed" directions are generally chosen as the direction of the Earth's spin axis and the perpendicular direction to the point of vernal equinox. These directions, however, are not exactly fixed owing to precession effects. The precession of the Earth's axis should thus be known if we are to introduce the appropriate corrections, but remember that in the final analysis the precession is also determined from observations of stars. The entire complex of problems is therefore solved simultaneously, and one has to consider the resultant effect of precession, galactic rotation, and Sun's motion, taking the sum of the peculiar components of the proper motions of star to be zero. The accuracy of the solution improves as we choose a progressively more uniform distribution of stars over the celestial sphere and take a sufficiently large sample. The proper motions determined in this way are called "absolute" in fundamental astrometry.

In photographic astromery the relative proper motions can be measured with high accuracy; these are motions relative to a group of reference stars on the photographic plate, the sum of whose proper motions is taken equal to zero. The conversion from relative to absolute proper motions is carried out using the results of fundamental astrometry which, as indicated above, depend on the value of the precession constant used. The relative proper motion is free from the effect of galactic rotation, except for terms of second and higher orders which are significant only for the distant stars. The linear terms are proportional to the distance, so that their angular equivalents remain constant and drop out when we take the difference between the measured and the reference stars.

Astrographic techniques can be applied to tag the proper motions of stars to extragalactic objects — galaxies whose distances are so large that

their direction from the Sun virtually remains unchanged. Calculations show that taking 300 km/sec for the tangential velocity of the galaxies, we obtain even for the nearest of them proper motions of 0".0001, which is one order of magnitude less than the current accuracy of proper motion determinations. The proper motions of stars relative to the galaxies may therefore be regarded as "absolute." These proper motions are independent of precession.

§138. Secular variation of proper motion, parallax, and radial velocity of a star due to its motion in space. Let μ, v_r, and p be the proper motion, the radial velocity, and the parallax of a star. The star is assumed to move in space with a constant velocity v over a certain length of time. In Figure 166 the two directions to the star cross its path at angles θ and θ' at distances r and r' from the observer O. During the time t, the star traverses the path tv. The angle between the two directions corresponding to this time interval is σ.

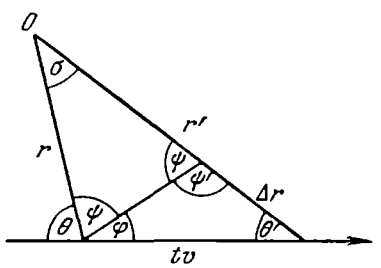


FIGURE 166

In the triangle of Figure 166

$$\frac{tv}{kr}\sin 1'' = \frac{\sin\sigma}{\sin\theta'},$$

where r is in parsecs, t is in years, v in km/sec; $k = 4.738$ (see §137).

Since $\theta' = \theta - \sigma$, we have

$$\frac{tv}{kr}\sin 1'' = \frac{\sin\sigma}{\sin\theta\cos\sigma - \cos\theta\sin\sigma} = \frac{\operatorname{tg}\sigma}{\sin\theta - \cos\theta\operatorname{tg}\sigma}.$$

The projections of the spatial velocity on the line of sight and the plane perpendicular to the line of sight at the starting time $t = 0$ are $v_r = v\cos\theta$ and $v_t = v\sin\theta$; the proper motion is $\mu = \frac{v_t}{kr}$, where r is in parsecs. Inserting $\sin\theta$ and $\cos\theta$ from these expressions in the above equation, we get

$$\operatorname{tg}\sigma = \frac{tkr\mu\sin 1''}{kr + tv_r\sin 1''}$$

or, since the angle σ is small,

$$\sigma = \frac{t\mu}{1 + \frac{tv_r\sin 1''}{kr}},$$

whence to fair approximation

$$\sigma = t\mu - t^2\mu v_r p \cdot 1.024 \cdot 10^{-6} \qquad (7)$$

Thus, when calculating the position of a star on a celestial sphere as a function of its proper motion, we should consider a quadratic term in time.

On the other hand, the proper motion, strictly speaking, is not $\frac{\sigma}{t}$, i.e., the angle between the two directions divided by the time.

The proper motion of a star at any time t is obtained by differentiating σ with respect to time, thus:

$$\frac{d\sigma}{dt} = \mu' = \mu - t\mu v_r p \cdot 2.05 \cdot 10^{-6} \qquad (8)$$

As an example consider Barnard's star, with $\mu = 10''.27$, $v_r = -111$ km/sec. and $p = 0''.55$. Then

$$\sigma = 10''.27t + t^2 \cdot 0''.00063.$$

If we ignore the quadratic term, the error in the star's position will reach $6''.3$ in 100 years.

The proper motion of Barnard's star will change by $+0''.126$ in 100 years.

Let us now derive equations for the secular variation of the parallax and the radial velocity. In Figure 166 we lay the segment r (equal to the initial distance from the observer) off the side r of the triangle. The distance increases by $\Delta r = r' - r$. In the new triangle with the sides Δr and r, the angles φ and $\psi' = 180 - \psi$ opposite these sides are

$$\varphi = 90^\circ - \left(\frac{\sigma}{2} - 0\right) \quad \text{and} \quad \psi = 90^\circ - \frac{\sigma}{2}$$

Therefore

$$\frac{r \sin 1''}{\Delta r} = \frac{\cos\frac{\sigma}{2}}{\cos\left(\frac{\sigma}{2} - 0\right)} = \frac{1}{\cos 0 + \operatorname{tg}\frac{\sigma}{2}\sin 0}$$

where Δr is again in parsecs, and t in years.

As before, we readily obtain

$$\Delta r = \frac{tv_r}{k} \sin 1'' + \frac{t^2}{2} \mu^2 r \sin^2 1''$$

For the difference of parallaxes we have

$$\Delta r = r' - r = \frac{1}{p'} - \frac{1}{p} = \frac{p - p'}{pp'},$$

Whence, dropping the quadratic term, we obtain to fair approximation

$$p' = p - tv_r p^2 \cdot 1.024 \cdot 10^{-6} \qquad (9)$$

The radial velocity of a star at the time t is found by differentiation of Δr with respect to time:

$$\frac{d\,\Delta r}{dt} = v_r' \frac{\text{parsec}}{\text{year}} = \frac{v_r \sin 1''}{k} + t\mu^2 r \sin^2 1''.$$

To express the result in km/sec, this equality should be multiplied by $\frac{k}{\sin 1''}$. Finally,

$$v_r' = v_r + \frac{t\mu^2}{p} \cdot 2.30 \cdot 10^{-5}. \qquad (10)$$

For Barnard's star the change in parallax and radial velocity in 100 years, calculated from (9) and (10), is

$$p' - p = +0''.0038 \quad \text{and} \quad v_r' - v_r = +1.2\,\text{km/sec}.$$

Werner (AN, B281, p. 221, 1954) tried to determine the kinematic parallax of the star Groombridge 1830 from the change in proper motion using Eq. (8), i.e.,

$$p = -\frac{(\mu' - \mu) \cdot 10^4}{t\mu v_r \cdot 2.05}.$$

Here $\mu = 7''.04$, $v_r = -98$ km/sec, $\mu' - \mu = +0''.019$, and $t = 100$ years. He found $p = 0''.14$, which fits the trigonometric parallax of the star, $0''.11$.*

§139. Statistical relations between the mean spatial velocity of a star and its radial and tangential components. We start with the assumption that, whatever the actual distribution of the residual velocities of stars, it is constant throughout the space. In practice, this is true at least in the immediate neighborhood of the Sun.

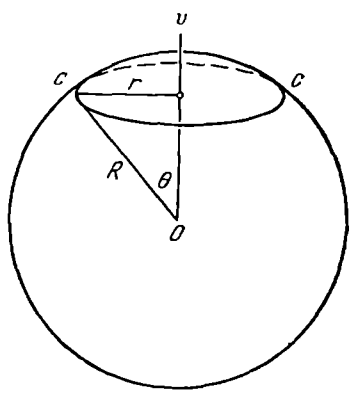


FIGURE 167

Consider some spatial velocity v, which has equal magnitude and direction at different points near the Sun. Its components v_r and v_t depend on the direction of the line of sight:

$$v_r = v\cos\theta \quad \text{and} \quad v_t = v\sin\theta,$$

where θ is the angle between the line of sight and the velocity vector v. The different projections are equiprobable, however. Indeed, the ray OC

* Werner erroneously used a factor of 1.024 in Eq. (8), instead of 2.05, and actually found $p = 0''.27$.

(Figure 167) which makes a constant angle θ with the velocity vector Ov meets the celestial sphere along the circle CC of radius $\varrho = R \sin\theta$, where R is the radius of the sphere. Hence the number of different projections of v_r and v_t is proportional to $\sin\theta$. Therefore, the mean magnitude of these projections is

$$|\bar{v}_r| = \frac{\int_0^{\frac{\pi}{2}} v \cos\theta \sin\theta \, d\theta}{\int_0^{\frac{\pi}{2}} \sin\theta \, d\theta} = \frac{1}{2} v,$$

$$|\bar{v}_t| = \frac{\int_0^{\frac{\pi}{2}} v \sin\theta \sin\theta \, d\theta}{\int_0^{\frac{\pi}{2}} \sin\theta \, d\theta} = \frac{\pi}{4} v.$$

Since this result is applicable to any spatial velocity and its projections, we have on the average for all velocities

$$|\bar{v}_r| = \frac{1}{2}\bar{v} \quad \text{and} \quad |\bar{v}_t| = \frac{\pi}{4}\bar{v}. \tag{11}$$

We thus have the following results:

1. The mean magnitude of the projection of a star's velocity on some axis is equal to half the mean spatial velocity of the star.
2. The mean magnitude of the projection of a star's velocity on some plane is equal to $\frac{\pi}{4}$ times the mean spatial velocity of the star.

These results were derived without assuming spherical distribution of spatial velocities. Any velocity distribution may be assumed, provided it is applicable in the entire space.

From Eq. (11) we have

$$|\bar{v}_t| = \frac{\pi}{2}|\bar{v}_r|. \tag{12}$$

§140. Proper motion components υ and τ. The observed proper motions of stars, like their radial velocities, include the projections of the Sun's reflected motion.

Consider a certain point, called the apex, on the celestial sphere (Figure 168), toward which the Sun moves relative to the nearby stars. The equatorial coordinates of the apex are $A_\odot$ and $D_\odot$. The antipodal point on the celestial sphere is the antapex. The position of any star with the coordinates α, δ is marked by the letter S. Its proper motion is described by the vector μ. The projections of this motion on the direction star—antapex and at right angles to this direction are the υ and τ components. Clearly the υ component includes a reflected motion of the Sun, i.e., the parallactic motion of the star υ_p. The component τ is free from this parallactic motion and corresponds only to the projection of the star's peculiar motion. The parallactic component also enters the radial velocity of the star.

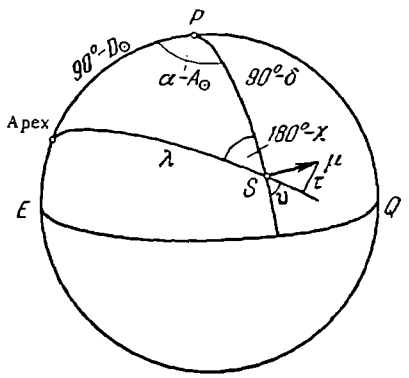


FIGURE 168

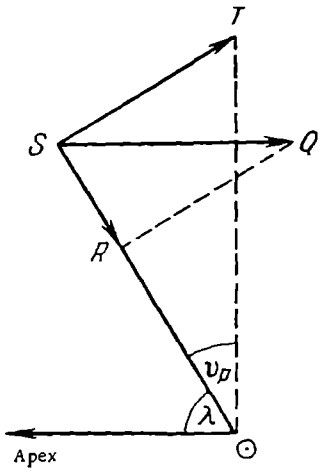


FIGURE 169

To change over from the proper motions $\mu_x = \mu_\alpha \cos\delta$ and $\mu_y = \mu_\delta$ to the υ and τ components, we use the position angle χ of the parallactic displacement of the star toward the antapex:

$$\left.\begin{aligned} \upsilon &= \mu_x \sin\chi + \mu_y \cos\chi, \\ \tau &= -\mu_x \cos\chi + \mu_y \sin\chi. \end{aligned}\right\} \qquad (13)$$

From the polar triangle pole—star—apex, we have using the standard relations of spherical trigonometry

$$\begin{aligned} \cos\lambda &= \sin\delta \sin D_\odot + \cos\delta \cos D_\odot \cos(\alpha - A_\odot), \\ \sin\lambda \cos\chi &= -\cos\delta \sin D_\odot + \\ &\quad + \sin\delta \cos D_\odot \cos(\alpha - A_\odot), \\ \sin\lambda \sin\chi &= \cos D_\odot \sin(\alpha - A_\odot), \end{aligned}$$

whence we find χ and λ, where λ is the angular distance of the star from the apex.

The parallactic displacement is a function of λ. If the direction to the star coincides with the direction to the apex or the antapex, the proper motion is free from the effect of Sun's motion. The radial velocity in this case includes the entire parallactic component. If the angular distance of the star from the apex is 90°, i.e., the directions to the star and to the apex are perpendicular to each other, the motion of the Sun is entirely included in the υ component and does not affect the radial velocity. In general, the parallactic component of proper motion is proportional to $\sin\lambda$ and that of the radial velocity to $\cos\lambda$. This is clear from Figure 169 where, as before, λ is then angle between the directions to the apex and to the star. The vector SQ is the reflected motion of the Sun in space. The projections of this vector on the line of sight and at right angles to it are

$$SR = SQ\cos\lambda \quad \text{and} \quad ST = SQ\sin\lambda. \qquad (14)$$

The vector ST subtends an angle v_p. For $\lambda = 90°$, $ST = SQ$ and the angle v_p reaches its maximum value, q. Introducing the distance r from the Sun to the star, we get

$$ST = r \sin v_p \quad \text{and} \quad SQ = r \sin q.$$

Inserting in the second equation in (14), we obtain

$$\sin v_p = \sin q \sin \lambda,$$

or, the angles being small (we express them in seconds of arc),

$$v_p'' = q'' \sin \lambda. \tag{15}$$

§141. Average secular and average annual parallaxes of stars. For each star with known proper motion we can find the v component from Eq. (13). Taking the average value $\bar{v}$ over a large sample of stars, we find that the peculiar component of the proper motions vanishes, since the velocity distribution of stars is on the whole symmetric. Although the stars which have large velocities relative to the Sun show definite asymmetry of motions (see §148), they can be excluded from analysis. Certain parts of the sky show local streams of stars which may distort the overall symmetry. Finally, the differential galactic rotation produces a biased effect, which is different in different parts of the sky (see §149).

All these factors force us to increase the star sample and to ensure maximum uniformity of distribution over the two celestial hemispheres. If this is done, we can safely say that $\bar{v}$ contains only the parallactic component of stellar motions.

The parallactic displacement of stars in itself is a poor characteristic of stellar distances, as it depends on the angle λ. One therefore generally calculates the parameter $\bar{q}$, which is independent of the star's angular distance from the Sun's apex.

Each star gives the equation

$$v = v_p + v',$$

where v' is the peculiar component. According to Eq. (15), this can be written in the form

$$v = q \sin \lambda + v'.$$

Solving by the least squares method, we obtain the most probable value of $\bar{q}$:

$$\bar{q} = \frac{\sum (v - v') \sin \lambda}{\sum \sin^2 \lambda} = \frac{\sum v \sin \lambda}{\sum \sin^2 \lambda}, \tag{16}$$

since

$$\frac{\sum v' \sin \lambda}{\sum \sin^2 \lambda} = 0.$$

$\bar{q}$ is called the mean secular parallax and is sometimes designated $\left(\frac{\bar{h}}{\varrho}\right)$, where h is the distance traversed by the Sun in one year, ϱ is the distance

to the star. In the notation of the previous section, this corresponds to SQ and r. The mean secular parallax is thus the angle which measures the mean parallactic motion of the selected group of stars referred to a direction at right angles to the direction to the apex.

Secular parallaxes characterize the average distances of stars grouped according to their apparent magnitudes and possibly according to their spectral types. In these groups the dispersion of the individual distances is reduced. Moreover, the stars are arranged according to galactic zones, since the average parallaxes of stars of a given magnitude depend on the galactic latitude: the luminosity function of the stellar population varies as we move from high latitudes toward the galactic plane.

Photographic observations give a large number of proper motions of stars of various apparent magnitudes, up to the very faintest, in small sky areas. Since within each small area the angles χ and λ are fairly constant, we get

$$\bar{v} = \bar{\mu}_x \sin\chi + \bar{\mu}_y \cos\chi = \bar{v}_p = \bar{q}\sin\lambda.$$

For numerous areas, distributed with fair uniformity over the sky, we get

$$\bar{q} = \frac{\sum v_p \sin\lambda}{\sum \sin^2\lambda}. \tag{17}$$

To change over from secular to average annual parallaxes $\bar{p}$, we require the velocity of the Sun $v_\odot$ relative to a given group of stars. Then

$$\bar{p} = \frac{4.74\cdot\bar{q}}{v_\odot}. \tag{18}$$

Taking 19.5 km/sec for the standard velocity of the Sun, we find that the annual parallax is approximately 1/4 of the secular parallax. The distances are thus determined with higher accuracy. Moreover, the proper motions from which the secular parallaxes are calculated are measured over periods of a few decades. The probable error of the secular parallaxes is thus about ±0".001. Hence it follows that the secular parallaxes give the distances to stars which are ten times farther than the stars with measurable trigonometric parallaxes. Trigonometric parallaxes are applicable to distances not exceeding 100—200 pc, and even then in few cases only.

Note that given the average parallax $\bar{p}$, we cannot calculate the average distance $\bar{r}$ as the reciprocal, since on account of the spread of the distances in each group

$$\bar{p}\cdot\bar{r} = \bar{p}\left(\frac{\overline{1}}{p}\right) \neq 1.$$

Special investigations show (see, e.g., Bull. Pulkovo Obs., No. 138, pp. 27—28, 1947) that the product $\bar{p}\cdot\bar{r}$ is approximately equal to two, if the stars are grouped only according to their apparent magnitudes. Thus, in order to obtain the average distance in parsecs, the reciprocal of the average parallax should be multiplied by a factor close to 2.

The average parallaxes of various groups of stars can also be found from the τ components. This requires knowledge of the peculiar radial

velocities of the stars. If we assume that the average absolute values of the τ components, expressed in angular units, correspond to the average absolute radial velocities expressed in km/sec, we get

$$\bar{p} = 4.74 \frac{|\bar{\tau}|}{|\bar{v}_r|}. \tag{19}$$

Remembering that the distribution of the spatial velocities of stars is ellipsoidal (see §143), we calculate the average radial velocity corresponding to a given direction of the τ component. If the star sample is uniformly distributed over the celestial sphere, this is not absolutely essential.

The uncertainty in the determination of the average parallaxes of stars of various apparent magnitudes from the τ components is due to the fact that the radial velocities are known only for the bright stars. For the faint stars, whose spectral types are generally unknown, we therefore use the radial velocities determined for the bright stars (see Oort, Bull. Astr. Inst. Netherl., Vol. VIII, No. 290, 1936).

Both methods for the determination of the average annual parallaxes — from the v and τ components — are used in stellar astronomy. Their application entails extensive catalogues of photographic and visual observations of proper motions. The most comprehensive and exact determination of the average parallaxes was carried out by Binnendijk in 1943 using photographic catalogues of proper motions in Kapteyn's selected areas (see §145) according to Pulkovo and Radcliffe observations. According to Binnendijk, the average annual parallaxes of stars up to m-th photographic magnitude are given by the equation

$$\lg \bar{p}(m) = \lg \bar{p}(12.0) - 0.115\,(m - 12.0),$$

where $\lg \bar{p}(12.0)$ takes the following values depending on the galactic latitude:

b	0°	10°	20°	30°	40°
$\lg \bar{p}(12.0)$	7.262	7.311	7.382	7.452	7.516
b	50°	60°	70°	80°	90°
$\lg \bar{p}(12.0)$	7.563	7.596	7.618	7.629	7.630

Table 1 lists the average parallaxes $\bar{p}$ calculated from Binnendijk's equation (in units of 0".0001).

TABLE 1

b \ m_{pg}	9.0	10.0	11.0	12.0	13.0	14.0	15.0
0°	40	31	24	18	14	11	8
±10	45	35	27	20	16	12	9
20	53	41	31	24	19	14	11
30	63	48	37	28	22	17	13
40	73	56	43	33	25	19	15
50	81	62	48	37	28	22	17
60	87	67	51	40	30	23	18
70	92	70	54	42	32	24	19
80	94	72	55	43	33	25	19
±90	94	72	56	43	33	25	19

To find the secular parallaxes $\bar{\pi}$, we should multiply the tabulated values by 4.11, assuming $v_\odot = 19.5$ km/sec.

§142. Determination of the Sun's apex from proper motions. In §137 we derived Eqs. (6), the first two of which related the proper motion components μ_x and μ_y to the projections $\dot{x}, \dot{y}, \dot{z}$ of the spatial velocity in a rectangular coordinate system oriented in a certain way.

The proper motions contain a parallactic and a peculiar component, and the right-hand sides of the equations can therefore be written as a sum of the projections $(-\dot{x}_\odot, -\dot{y}_\odot, -\dot{z}_\odot)$ of the reflected motion of the Sun and the individual velocity of the star. The latter is of no importance for our purposes and can therefore be written in a general form as μ_x' and μ_y'.

For each star we thus have the two equations

$$\left.\begin{aligned} \mu_x &= \frac{\dot{x}_\odot \sin\alpha - \dot{y}_\odot \cos\alpha}{kr} + \mu_x', \\ \mu_y &= \frac{\dot{x}_\odot \cos\alpha \sin\delta + \dot{y}_\odot \sin\alpha \sin\delta - \dot{z}_\odot \cos\delta}{kr} + \mu_y'. \end{aligned}\right\} \qquad (20)$$

Apart from the sought components $\dot{x}_\odot, \dot{y}_\odot, \dot{z}_\odot$, the peculiar components μ_x', μ_y' are also unknown. Given a sufficient number of stars distributed fairly uniformly over the celestial sphere, the peculiar motions in all probability mutually cancel. The distances r from the Sun to the stars are generally unknown, especially for the weak stars. We therefore have to use some average distance $\bar{r}$ for all the relevant stars and find by the least squares method the unknowns $\frac{\dot{x}_\odot}{\bar{r}}, \frac{\dot{y}_\odot}{\bar{r}}, \frac{\dot{z}_\odot}{\bar{r}}$.

To reduce the dispersion of stellar distances, the stars are divided into groups according to apparent magnitudes. A certain factor can be introduced to allow for the difference in the average distances of stars at different galactic latitudes using, e.g., the ratio of the average secular parallaxes.

In photographic proper motion determinations, when each photographed area contains a large number of stars, one generally calculates the average $\bar{\mu}_x$ and $\bar{\mu}_y$ for stars of various apparent magnitudes in a given area, and the α and δ of the individual stars can be replaced by the coordinates α_c and δ_c of the area center. The peculiar motions are on the average set equal to zero. Eqs. (20) are then written in the form

$$\bar{\mu}_x = \frac{\dot{x}_\odot \sin\alpha_c - \dot{y}_\odot \cos\alpha_c}{k\bar{r}}, \quad \bar{\mu}_y = \frac{\dot{x}_\odot \cos\alpha_c \sin\delta_c + \dot{y}_\odot \sin\alpha_c \sin\delta_c - \dot{z}_\odot \cos\delta_c}{k\bar{r}}. \qquad (21)$$

To find the projections of the Sun's velocity, a set of these equations drawn up for numerous areas is solved by the least squares method. Combining the two systems of normal equations obtained from (21) for μ_x and μ_y, we get

$$\begin{aligned} \dot{\zeta}_\odot \sum (1 - \cos^2\alpha \cos^2\delta) - \dot{\eta}_\odot \sum \sin\alpha \cos\alpha \cos^2\delta - \dot{\vartheta}_\odot \sum \cos\alpha \sin\delta \cos\delta = \\ = \sum \bar{\mu}_x \sin\alpha + \sum \bar{\mu}_y \cos\alpha \sin\delta, \\ -\dot{\zeta}_\odot \sum \sin\alpha \cos\alpha \cos^2\delta + \dot{\eta}_\odot \sum (1 - \sin^2\alpha \cos^2\delta) - \dot{\vartheta}_\odot \sum \sin\alpha \sin\delta \cos\delta = \\ = -\sum \bar{\mu}_x \cos\alpha + \sum \bar{\mu}_y \sin\alpha \sin\delta, \\ -\dot{\zeta}_\odot \sum \cos\alpha \sin\delta \cos\delta - \dot{\eta}_\odot \sum \sin\alpha \sin\delta \cos\delta + \dot{\vartheta}_\odot \sum \cos^2\delta = -\sum \bar{\mu}_y \cos\delta. \end{aligned}$$

Here for brevity $\dot{\zeta}_\odot, \dot{\eta}_\odot, \dot{\vartheta}_\odot$ stand for $\frac{\dot{x}_\odot}{k\bar{r}}$, $\frac{\dot{y}_\odot}{k\bar{r}}$, $\frac{\dot{z}_\odot}{k\bar{r}}$, respectively, and the subscript c of α and δ has been omitted.

In general, Eqs. (20) and (21) contain as unknowns the precession corrections and galactic rotation terms. If the precession and the galactic rotation are not required, the proper motions are first corrected for these effects.

Having found $\dot{\zeta}_\odot, \dot{\eta}_\odot$ and $\dot{\vartheta}_\odot$, we insert the results in Eqs. (5) dividing both sides by $k\bar{r}$. For a star situated precisely at the Sun's apex at an average distance $\bar{r}$ with zero peculiar motion, Eqs. (5) take the form

$$\left.\begin{aligned} \dot{\zeta}_\odot &= \frac{v_\odot}{k\bar{r}} \cos A_\odot \cos D_\odot, \\ \dot{\eta}_\odot &= \frac{v_\odot}{k\bar{r}} \sin A_\odot \cos D_\odot, \\ \dot{\vartheta}_\odot &= \frac{v_\odot}{k\bar{r}} \sin D_\odot, \end{aligned}\right\} \tag{22}$$

where $A_\odot$ and $D_\odot$ are the right ascension and the declination of the Sun's apex, $v_\odot$ is the total velocity of the Sun in km/sec.

From Eq. (22) we readily get

$$\left.\begin{aligned} \operatorname{tg} A_\odot &= \frac{\dot{\eta}_\odot}{\dot{\zeta}_\odot}; \quad \operatorname{tg} D_\odot = \frac{\dot{\vartheta}_\odot}{\sqrt{\dot{\zeta}_\odot^2 + \dot{\eta}_\odot^2}}, \\ \frac{v_\odot}{k\bar{r}} &= \sqrt{\dot{\zeta}_\odot^2 + \dot{\eta}_\odot^2 + \dot{\vartheta}_\odot^2}. \end{aligned}\right\} \tag{23}$$

Note that $\dot{\zeta}_\odot, \dot{\eta}_\odot, \dot{\vartheta}_\odot$, like $\frac{v_\odot}{k\bar{r}}$, are expressed in seconds of arc per year. Hence $\frac{v_\odot}{k\bar{r}}$ is in fact the secular parallax of the given group of stars. The coefficient $k = 4.738$ (see §137). To find the velocity of the Sun in km/sec, the average distance $\bar{r}$ is required (in parsecs).

The fundamental equations (20) were proposed in 1859 independently by two authors, M. A. Koval'skii (Recherches astronomiques de l'observatoire de Kasan, No. 1) and G. Airy (M. N. R. Astronom. Soc., Vol. 19, 175). This method for the determination of the elements of Sun's motion is thus appropriately called the Airy-Koval'skii method (see D. Ya. Martynov.—Astron. Zh., 27, No. 3:169. 1950).

The coordinates of the solar apex have been repeatedly determined from the proper motions of various star groups. For the bright stars approximately up to 9th apparent magnitude, the apex coordinates are found to be close to

$$A_\odot = 270°, \quad D_\odot = +30° \qquad (1900.0).$$

We shall see in the following that the radial velocities of the bright stars give on the average the same results, and these apex coordinates are therefore used as the standard. In the galactic system of coordinates we have

$$l_\odot = 24°, \qquad b_\odot = +22°.$$

The fainter stars give a different apex for the Sun. Thus according to Pulkovo determinations for stars between photographic magnitudes 9 and 15 measured in Kapteyn's areas

$$A_\odot = 285° \pm 2°, \quad D_\odot = +36° \pm 1°.$$

The difference is possibly due to the higher percentage content of the dwarfs among the faint stars, as their velocities relative to the Sun are high.

Stars with large proper motions increase the right ascension of the apex by 15—20° and the declination by 5—10° compared to the standard values.

On the other hand, McCormick Observatory data indicate that the position of the apex depends on the system of absolute proper motions used. The apex coordinates derived from the proper motions of the faint stars up to 13th magnitude in the GC system were, as in Pulkovo,

$$A_\odot = 285° \pm 3°, \quad D_\odot = +36° \pm 2°,$$

whereas calculations using the FK3 absolute system gave results very close to the standard apex

$$A_\odot = 274° \pm 2°, \quad D_\odot = +30° \pm 1°.$$

The results are apparently influenced by the biased errors of the fundamental catalogues.

The apex position determined from proper motions showed no dependence on the spectral type of the stars.

It should be emphasized, however, that in relation to various special groups of celestial objects, e.g., certain types of variables, globular clusters, galaxies, the coordinates of the Sun's apex and its velocity may greatly deviate from the standard values. This is particularly pronounced in the study of radial velocities, which can be determined for the relatively distant objects too.

§143. Determination of the elements of the Sun's motion from radial velocities. Consider the third equation in (6), in which we isolate the peculiar component of radial velocity v'_r. Then

$$v_r = -\dot{x}_\odot \cos\alpha\cos\delta - \dot{y}_\odot \sin\alpha\cos\delta - \dot{z}_\odot \sin\delta + v'_r, \tag{24}$$

where $-\dot{x}_\odot$, $-\dot{y}_\odot$, $-\dot{z}_\odot$ are the projections of the reflected motion of the Sun.

Given a sufficient number of these equations, the least squares method can be applied to find the sought projections of the solar motion assuming that the peculiar components mutually cancel. Then, by analogy with Eqs. (23), we have

$$\left.\begin{aligned} \operatorname{tg} A_\odot &= \frac{\dot{y}_\odot}{\dot{x}_\odot}; \quad \operatorname{tg} D_\odot = \frac{\dot{z}_\odot}{\sqrt{\dot{x}_\odot^2 + \dot{y}_\odot^2}}; \\ v_\odot &= \sqrt{\dot{x}_\odot^2 + \dot{y}_\odot^2 + \dot{z}_\odot^2}. \end{aligned}\right\} \tag{25}$$

Here the velocity of the Sun $v_\odot$ is in km/sec.

The coordinates of the Sun's apex according to various determination of the radial velocities of bright stars up to apparent magnitude 5—6 gave the standard value

$$A_\odot = 270°, \quad D_\odot = +30° \qquad (1900.0).$$

The velocity of the Sun relative to these stars was found to be

$$v_\odot = 19.5 \text{ km/sec},$$

which was also adopted as the standard value.

However, the velocity of the Sun and the apex coordinates in relation to special groups of celestial objects show marked deviation from the standard.

Thus, in relation to the long-period cepheids, we have

$$A_\odot = 287°, \quad D_\odot = +40°, \quad v_\odot = 19 \text{ km/sec} \quad \text{(Parenago, 1946)}.$$

In relation to the short-period cepheids

$$A_\odot = 303°, \quad D_\odot = +41°, \quad v_\odot = 140 \text{ km/sec} \quad \text{(Pavlovskaya, 1953)}.$$

The globular cluster give

$$A_\odot = 301°, \quad D_\odot = +48°, \quad v_\odot = 168 \text{ km/sec} \quad \text{(Mayall, 1946)}.$$

For extragalactic nebulae

$$A_\odot = 345°, \quad D_\odot = +50°, \quad v_\odot = 290 \text{ km/sec} \quad \text{(Humason, 1955)}.$$

No clear dependence on spectral characteristics of stars was observed. The right ascension of the Sun's apex somewhat increases on passing from the early main sequence types to late types. This is accompanied by a certain increase (approximately by a factor of 1.5) in the velocity $v_\odot$ relative to the dwarfs compared to the velocity relative to the giants.

In Parenago's opinion, all the kinematic characteristics of the main sequence stars, the parameters of the Sun's motion included, markedly change in the spectral type F. Thus, the coordinates of the Sun's apex and its velocity relative to stars of spectral types from B to F are almost constant, then they change rapidly ($v_\odot$ and $A_\odot$ increase, $D_\odot$ decreases) in spectral type F, and remain without any further substantial change in types G, K, and M.

§144. The K effect. In early works on the determination of Sun's motion from radial velocities it was noted that the introduction of a constant term in Eq. (24) in some cases ensures a better fit with observations. This term is known as the K term, and the corresponding phenomenon as the K effect,

$$v_r = K - \dot{x}_\odot \cos\alpha\cos\delta - \dot{y}_\odot \sin\alpha\cos\delta - \dot{z}_\odot \sin\delta + v'_r. \qquad (26)$$

The K term was found to depend on the spectral type of the star. It is maximum for stars of types O and B, decreases markedly for A stars and is almost zero for late-type stars.

Table 2 lists the values of the K term for various spectral types according to the determinations of different authors. The errors in these determinations are of the order of a few tenths of a kilometer per sec.

The nature of the K effect is not entirely clear. Literally, the K effect signifies that all the nearby O—B stars move away from the Sun in all directions. Systematic errors in the determination of spectral line wavelengths used in radial velocity measurements or hypothetical motions in the atmospheric layers of the B stars cannot fully account for this effect.

TABLE 2

Spectrum	K-term, km/sec	Number of stars
O—B	+4.8	929
A	+0.7	1212
F	−0.3	722
G	−0.7	677
K	0.0	1805
M	+0.4	456

The K effect is partly explained by the gravitational red shift of the spectral lines in the field of the massive stars according to the general theory of relativity. The calculations of Plaskett and Pearce give +1.3 km/sec for the gravitational red shift of B stars, which is one fourth of the total K effect.

B stars often form groups with common motion, and the constant radial velocity component is possibly associated with the group motion of at least some of the stars. A local stream of B stars is observed in the constellations of Scorpius—Centaurus in the Southern Hemisphere. Note that according to the observations of some authors the K effect is observed only for the bright B stars. Thus, Nordstrom found for 353 stars brighter than 6 mag. $K = +4.2 \pm 0.6$ km/sec, and for 379 faint stars $K = -1.8 \pm 0.9$ km/sec.

§145. Kapteyn's "selected areas". We do not require complete information on all the stars in the sky for purposes of statistical studies in stellar astronomy. This would involve enormous expenditure of effort and could hardly be carried to completion by all the observatories in the worlds.

In 1906 the Dutch astronomer Kapteyn proposed a program for the investigation of various characteristics of stars, up to the very faintest, in selected areas of the sky. This selective approach, not unlike William Herschel's method of "scoops," provides statistically valid information on the structure of the Galaxy and its dynamics. Kapteyn's program of selected areas comprises two parts: systematic, in which 206 areas uniformly distributed over the sky are examined, and special, encompassing 46 areas, mainly lying in the Milky Way. The first step is to carry out the systematic part of the program.

TABLE 3

δ	Number of stars
±90°	2
±75	12
±60	24
±45	48
±30	48
±15	48
0	24

The frequency of declinations of the centers of the 206 areas is given in Table 3. In right ascension the areas are distributed at one-hour or larger intervals, depending on declination. The size of the areas is not precisely defined: it depends on the instrument used and the aim of the observations. The field size is generally between one and a few square degrees.

For the stars in the selected areas one determines the position, the photographic, visual, and photovisual magnitudes, color, spectrum, proper motion, radial velocity, and other characteristics. The limiting stellar magnitude clearly differs according to the instrumental methods used in the determination of the various parameters.

Over the last 50 years Kapteyn's program underwent some changes. For example, the determination of trigonometric parallaxes was dropped, since the distances of the great majority of stars were found to lie beyond the

accuracy of trigonometric measurements. New problems arose, involving determination of photoelectric stellar magnitudes, absolute stellar magnitudes, star light polarization, etc.

Numerous observatories in various countries take part in Kapteyn's program, which currently comprises the following stages.

a) Catalogue of positions and stellar magnitudes. A catalogue listing the positions (up to 1^s in α and 1" in δ) and the photographic magnitudes of all the stars up to 16 mag. on the average was compiled for the 206 systematic areas. Field size 40'×40', 60'×60', and 80'×80' for low, medium, and high galactic latitudes, respectively. The equatorial coordinates of the center of each area are given. The work was carried out in Groningen (Holland) using the Harvard plates and published in 1918—1924 in three volumes.

An analogous catalogue for the 46 special areas was published in 1952 in two volumes.

b) Photographic, visual, and photovisual magnitudes; color indices and spectra. The photographic magnitudes of stars up to 18—19 mag. were determined at the Mount Wilson Observatory for 39 of the systematic areas from the North Pole to declination circle −15°. Field size 15'×15' at galactic latitudes below 40° and 20'×20' for higher latitudes. For areas at latitudes below 40° additional stars are given, which increase the effective field size almost to 20'×20'. This catalogue was published in 1930.

Another project of the Mount Wilson Observatory is the publication of a catalogue of photovisual stellar magnitudes to 14—16 mag. for the same areas.

For the southern zones from −15 to −90° the photographic magnitudes of stars brighter than $10^m.5$ were determined at the Observatory of Leiden from photographs taken in Johannesburg (South Africa). Field size 16 sq. degrees. The catalogue was published in 1957. The survey is continued with the object of covering all stars up to 14 mag. in these areas.

The Abastumani Observatory completed the determination of the color indices for stars from 10 to 13 mag. in 43 areas in zones from the North Pole to declination +45°. Each area was 2×2°. The catalogue was published in 1949. The project is now continued for the zone +30°.

Tikhov's method of the longitudinal spectrograph was applied at the Pulkovo Observatory to determine the color indices of the Bonner Durchmusterung (BD) stars which fall in Kapteyn's areas. Field diameter 4°. Catalogues for zones from +90 to +15° were published at the Pulkovo Observatory in 1937 and 1951. Catalogues for zones 0° and −15° are in preparation.

The Hamburg Observatory in Bergedorf published a catalogue of star spectra up to 13th photographic magnitude for all the 115 northern Kapteyn's areas, the equator included. Field size 3°.5×3°.5. The catalogue was published in 1935—1953 in five volumes. The photographic stellar magnitudes in this catalogue were measured at Groningen from Harvard plates.

In addition to the above projects, various observatories carried out determinations of stellar magnitudes in individual zones and areas. A complete list of publications and detailed information on current and future projects will be found in the reports of Committee 32 of the International Astronomical Union. Since 1961 they are published in the reports of Subcommittee 33_C.

c) Proper motions. The relative proper motions of thousands of stars up to 15 mag. in the northern areas were determined at the Radcliffe Observatory (Oxford) and at Pulkovo. Field size from 20'×20' to 80'×80'. The corresponding catalogues were published in 1934 and 1940.

Photographic observations of the first epoch have been completed in some other observatories, and second-epoch photographs are now being taken. Algiers, Helsinki, Cape and other observatories use old Carte du Ciel photographs coinciding with the selected areas for the determination of proper motions of stars up to 12 mag.

d) Radial velocities. The radial velocities of the brightest stars in the northern sky areas were determined in some observatories. The radial velocities of 467 stars brighter than 9 mag. in 139 areas from declination +90 to −15° were determined at the Mount Wilson Observatory and published on pp. 485—495 of the 1952 IAU Proceedings.

For areas 140—206 in the Southern Hemisphere, measurements of radial motions of the GC stars with proper motions over 0".1 are being carried out at the Cape Observatory.

Haute-Provence Observatory (France) has launched a project involving determination of the radial velocities of stars up to 10 mag. by a direct-sight objective prism in all the Northern Hemisphere areas. Field size 3°×4°.

Other characteristics of stars in the selected areas are also being determined. Thus, at Stockholm, the spectroscopic absolute magnitudes of stars up to 13.5 photographic mag. in 15 northern areas were determined. The results were published in 1951, 1955, and 1956.

At the Abastumani Observatory, the spectroscopic absolute magnitudes of 766 B—A stars brighter than 9 mag. in 44 Kapteyn's areas were determined. The results were published in 1958.

Variable stars, binary stars, and stars with large proper motions are being studied in Kapteyn's areas. Full bibliography is currently published in the reports of the IAU Committees.

The observational data collected in various observatories under Kapteyn's program served as the basis for numerous stellar-statistical studies pertaining to the structure and the dynamics of the Galaxy.

Note that new programs of galactic research were laid down in recent years following the rapid advances in various branches of astronomy. Thus, the Pulkovo Observatory devised a program of photographic observations of galaxies with the object of tagging the proper motions of stars up to 17 mag. in 300 areas all over the sky. The area size is about 4 sq. degrees. The program has been successfully launched on an international scale, in numerous observatories over the world.

The Lick Observatory has completed the set of first-epoch photographs for the entire sky area from the north pole to declination −23° on plates covering 36 sq. degrees each. These photographs will eventually be used to determine the absolute proper motions of stars up to 18.5 mag. relative to the weak galaxies.

Parenago (the Shternberg Institute) proposed a complex program for a number of areas of particular interest in the Milky Way. Numerous observatories participate in this program.

§146. The centroid. Residual star velocities. Let $P(x, y, z)$ be a point in the star's system. Consider a volume Q containing this point and a

sufficiently large number of stars n. The star velocities in this volume follow a certain distribution. We can calculate the average velocity

$$\frac{1}{n}\sum_{i=1}^{n} v_i, \tag{27}$$

where v_i is the velocity vector of the i-th star.

The average velocity, which is in fact the velocity of the geometrical center of the given system of stars, differs from the center of inertia velocity of the system in that it does not account for the difference in stellar masses.

Let us now make the volume Q progressively smaller. The average velocity (27) changes in the process. Formally the centroid velocity of $P(x, y, z)$ is the limit of the vector (27) as the volume Q goes to zero so that $n \to 1$. The limit must be independent of the actual manner in which the volume is contracted.

As the stellar system is a discrete structure and the stars all have different velocities, the mean (27) has no limit in the proper sense as Q is contracted to zero. As Q is made progressively smaller, the change in the vector (27) following the omission of each successive star from the volume becomes progressively more significant and the final (for $n = 1$) value of this vector depends on the particular star which remains, together with the point P, inside the infinitesimal volume Q.

Therefore, the concept of a centroid and centroid velocity defy rigorous definition. The following approximate definition can be advanced, however. The centroid velocity of $P(x, y, z)$ is the average velocity of the stars inside a sphere of radius r centered at $P(x, y, z)$. No exact value can be assigned to the radius r. It is approximately estimated in the following way. Consider an assembly of spheres of a fixed radius r tangent to the sphere Q. For each of these adjoining spheres we can calculate the average velocity of the contained stars. Let $\Delta\bar{v}$ be the largest difference between the average velocity vectors of the spherical volume Q and an adjoining spherical volume. The radius r should be decreased until the change in (27) associated with individual stars lying outside the sphere is small compared to $\Delta\bar{v}$. The radius r defined in this way in rotating stellar systems is generally small compared to the radius of the entire system, and yet the number of stars in the sphere Q will be large compared to unity.

If we consider two neighboring stars and the centroids coinciding with these stars, the difference in centroid velocities in general is much less than the difference in the velocities of the stars. The introduction of the concept of centroids thus imparts to the stellar system a measure of continuity.

Besides the general concept of a centroid, we are often dealing with special centroids. A special centroid is obtained if a certain subgroup with some common qualifying characteristic is considered out of all the stars inside the volume Q, such as stars of a given spectral type or stars with large proper motions. Special centroids clearly differ from the general centroid.

The difference between the velocity of a star and the centroid velocity at the point occupied by that star is called the residual velocity of the star;

for example, the velocity of the Sun — 19.5 km/sec toward the apex — is actually its residual velocity. Sometimes the term peculiar velocity is used for residual velocity.

The kinematic properties of a stellar system are conveniently studied by considering the centroid velocity at each point and the distribution function of the residual velocities at that point, instead of the total velocity distribution function of stars.

Since the velocities of nearby centroids are not much different from one another, the residual velocities of the nearest stars are in practice determined by adding to the observed velocity the residual velocity of the Sun or, equivalently, subtracting from the observed velocity the parallactic velocity of the star associated with the motion of the Sun toward its apex. To find the residual velocities of distant stars, we should additionally consider the difference between the centroid velocities of the Sun and the stars.

§147. Distribution of residual velocities. As we know, the velocities of molecules in a stationary gas show a Maxwellian distribution. The projections of the molecular velocities on any direction are distributed as

$$F(\gamma)\,d\gamma = \frac{N}{\sigma\sqrt{2\pi}}\,e^{-\frac{\gamma^2}{2\sigma^2}}\,d\gamma, \tag{28}$$

where σ is constant in all directions. If u, v, w are the rectangular velocity components of the molecules, the Maxwellian distribution of the three components is given by

$$F(u, v, w)\,du\,dv\,dw = \frac{N}{(2\pi)^{3/2}\sigma^3}\,e^{-\frac{1}{2\sigma^2}(u^2+v^2+w^2)}\,du\,dv\,dw. \tag{29}$$

In (29), the surfaces of constant density have the form

$$u^2 + v^2 + w^2 = \text{const},$$

i.e., these are spheres. The Maxwellian distribution is therefore a spherical distribution. This is a particular case of the general spherical distribution function

$$F(u, v, w)\,du\,dv\,dw = f(u^2 + v^2 + w^2)\,du\,dv\,dw. \tag{30}$$

In a spherical distribution, all the directions are equivalent.

Up to the beginning of the 20th century, it was generally assumed that the residual velocities of stars, by analogy with gas molecules, are distributed according to (29).

In 1904 Kapteyn established, however, that the star velocities have two preferred directions, which we called the v e r t i c e s. If we eliminate the velocity of the Sun, i.e., consider the residual velocities of stars, the coordinates of the two vertices are

$$\text{first vertex } \alpha = 91^\circ, \quad \delta = +13^\circ,$$
$$\text{second vertex } \alpha = 271^\circ, \quad \delta = -13^\circ.$$

The two vertices are thus at diametrically opposing points.

Proceeding from this result, Kapteyn advanced his two-stream hypothesis. The star streams move in countercurrents, one with velocity u_1

and the other with velocity u_2, and are by now thoroughly intermixed. The residual velocities of stars in each stream relative to that stream (considered as the centroid) show a Maxwellian distribution. The overall distribution of the residual velocities can thus be written in the form

$$F(u, v, w)\,du\,dv\,dw = \frac{1}{(2\pi)^{3/2}}\left[\frac{N_1}{\sigma_1^3}\,e^{-\frac{1}{2\sigma_1^2}[(u-u_1)^2+v^2+w^2]} + \frac{N_2}{\sigma_2}\,e^{-\frac{1}{2\sigma_2^2}[(u-u_2)^2+v^2+w^2]}\right] du\,dv\,dw.$$

Unlike the Maxwellian distribution, which is described by a single parameter σ, the distribution in Kapteyn's two-stream theory, requires seven parameters, $\frac{N_1}{N_2}$, σ_1, σ_2, u_1, u_2, and the two coordinates defining the direction of u_1 (and u_2). The theory of Kapteyn's streams was developed in detail by Eddington. It follows from Eddington's results that the stellar population in each volume of space contains representatives of both streams. Therefore, if the existence of two streams is assumed, we are led to the conclusion that the two streams are ideally mixed. This feature imparts a definite flavor of artificiality to the two-stream hypothesis.

A more natural point of view was advanced by K. Schwarzschild (1912). He simply suggested that the various directions are not equiprobable for the residual stellar velocities. The direction to the vertices of Kapteyn's streams is a preferred direction for some reason. The velocity projections on this direction are distributed according to (28), but since the velocities are greater in this direction, the parameter σ characterizing the dispersion of the velocities is also relatively large.

Thus, assuming a Maxwellian distribution but using different dispersions of the residual stellar velocities in the three mutually perpendicular directions, Schwarzschild proposed the following distribution for the rectangular components of the residual velocity vector of stars:

$$F(u, v, w)\,du\,dv\,dw = \frac{N}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3}\,e^{-\frac{1}{2}\left(\frac{u^2}{\sigma_1^2}+\frac{v^2}{\sigma_2^2}+\frac{w^2}{\sigma_3^2}\right)} du\,dv\,dw. \qquad (31)$$

In this distribution the surfaces of equal density are ellipsoids:

$$\frac{v^2}{\sigma_1^2}+\frac{v^2}{\sigma_2^2}+\frac{w^2}{\sigma_3^2} = \text{const.} \qquad (32)$$

Schwarzschild's distribution is therefore often called an ellipsoidal distribution. In fact, the general ellipsoidal distribution is

$$F(u, v, w)\,du\,dv\,dw = f\left(\frac{u^2}{\sigma_1^2}+\frac{v^2}{\sigma_2^2}+\frac{w^2}{\sigma_3^2}\right) du\,dv\,dw, \qquad (33)$$

so that Schwarzschild's distribution is a particular form of the general law.

Schwarzschild's distribution has six parameters, $\sigma_1, \sigma_2, \sigma_3$, two angular coordinates which specify the direction of the component u, and one angular coordinate specifying the direction of the component v.

Several methods have been proposed for the determination of these parameters from radial velocities of stars, their proper motions, and spatial motions (see, e.g., /15/), and these parameters were calculated from extensive statistical data.

TABLE 4

Star groups	N	L	B	$v_\odot$	σ_1	σ_2	σ_3	Author
B0—B5	40	18°	+22°	18	10	8	5	Parenago, 1949
B6—A2	53	12	+25	16	16	10	6	" "
A3—A8	61	6	+20	18	19	9	8	" "
A9—F1	110	11	+24	17	24	13	9	" "
dF2—dF4	180	21	+28	16	27	17	12	" "
dF5—dF7	177	39	+22	21	32	21	17	" "
dF8—dG2	188	32	+6	33	46	28	23	" "
dG3—dG7	113	33	+3	37	50	30	27	" "
dG8—dK2	100	30	+14	36	52	30	23	" "
dK3—dK5	121	46	+14	38	51	29	23	" "
dM	170	38	+20	26	46	26	22	" "
gA—gF	59	10	+20	27	27	14	10	" "
gG—gM	1226	31	+19	19	30	20	16	" "
sgF—sgK	113	45	+20	30	42	27	24	" "
cO—cM	173	23	+21	16	13	10	8	" "
Irregular and semiregular M-type variables	45	41	+27	30	38	23	19	" "
High-velocity giants	20	48	+11	106	100	60	50	" "
Short-period cepheids	132	49	+4	140	122	122	65	Pavlovskaya, 1953
Interstellar gas clouds	837	27	+23	18	7	7	5	Erleksova, 1952
Long-period cepheids	125	38	+12	19	12	9	5	Parenago, 1946
Open clusters	26	37	+21	22	15	12	6	Barkhatova, 1949
R and N stars	88	20	+32	23	30	28	38	Wilson, 1935 and Ikaunieks, 1952
Planetary nebulae	96	41	+13	31	45	35	20	Wirtu, 1922
Globular clusters	50	51	+8	168	110	—	—	Mayall, 1946
White dwarfs	50	32	0	45	45	28	20	Parenago, 1949
Subdwarfs	141	59	+3	136	125	96	64	" "
Mira Ceti stars, except those with periods of 150—200 days	86	44	+18	24	38	30	29	Kulikovskii, 1948
Mira Ceti stars with periods of 150—200 days	13	35	+7	120	91	45	36	" "

The results are on the whole fairly consistent and amount to the following.

1) The large axis of the velocity ellipsoid is directed toward the center of the Galaxy, the medium axis lies in the plane of the Galaxy at right angles to the direction to the center, and the small axis is parallel to the galactic axis of symmetry.

2) The ellipsoid axis ratio is 1:0.62:0.50. This quantity is a highly stable characteristic (individual deviations do not exceed a few percent) which hardly depends on the type of stars considered.

3) Whereas the ellipsoid axis ratio is highly stable for various statistical ensembles of stars, the actual length of these axes is greatly variable. This is evident from Table 4, borrowed from P. P. Parenago's "Course in Stellar Astronomy," which gives the results of determinations of the ellipsoid axes for various groups of stars and other objects.

In this table, L and B are the galactic longitude and latitude specifying the direction of the Sun's velocity relative to the centroid of stars being

considered, $v_\odot$ is the magnitude of this velocity in km/sec, σ_1, σ_2, σ_3 are the semiaxes of the velocity ellipsoid in km/sec.

The change in L, B, and $v_\odot$ for objects of different types shows that these objects have different centroid velocities. The reason for this will be considered in §148 and §149. The semiaxes of the velocity ellipsoid are also highly variable. A very significant feature is that as we move from early to late spectral types, the semiaxes of the velocity ellipsoid at first slowly increase between B0 and dF4, then rise rapidly between dF5 and dG7, and finally remain nearly constant after dG8.

Several hypotheses have been advanced to explain this phenomenon. According to Parenago (1951), the main sequence stars in fact fall into two subsequences. This division is of cosmogonic character. The stars of the first sequence (from O to F) form in the disk subsystems and have low residual velocities (25 km/sec on the average), whereas the stars of the second sequence (from F to M) form in the intermediate subsystems and have large residual velocities (60 km/sec on the average). These two sequences overlap in the F—G region, which produces the steep increase in residual velocities rather than a discontinuous jump.

The increase of the residual velocities between O and B, according to Agekyan (1950), is attributable to the interaction (gravitational and radiation pressure) of these high-luminosity stars with diffuse interstellar matter.

M. Schwarzschild and Spitzer (1951) advanced a hypothesis which associates the large residual velocities of the late-type stars with the gravitational interaction (ignoring radiation pressure) between these stars and the massive dust clouds ($m = 10^6\, m_\odot$).

The ellipsoidal velocity distribution is the most convenient approximation to the true residual velocity distribution of stars. It is based on an analogy with the Maxwellian distribution of the velocities of gas molecules. Boltzmann has proved rigorously that the Maxwellian distribution in a gas is established as a result of molecular interactions. Under normal conditions it takes a negligible fraction of a second to reach a Maxwellian distribution in a gas.

The interaction between the stars should also lead to a Maxwellian velocity distribution. In open clusters, where stellar ecnounters are fairly frequent, the actual velocity distribution should in fact be close to Maxwell's, and the small deviations from the Maxwellian distribution are associated with the inevitable escape from the clusters of stars with supercritical velocities. In the galactic field, on the other hand, a tremendous length of time, substantially exceeding the age of our Galaxy, is required for the establishment of Maxwellian velocity distribution. It will be shown in §158 that the ellipsoidal distribution is at variance with observations. Dynamic considerations only give an elliptic distribution of the projections of small residual velocities on the plane of the Galaxy,

$$F(u, v)\, du\, dv = f\left(\frac{u^2}{\sigma_1^2} + \frac{v^2}{\sigma_2^2}\right) du\, dv. \tag{34}$$

There are no indications that this distribution actually coincides with Schwarzschild's distribution (31).

The general problem of determining the distribution of the residual velocities from the observed radial velocity distribution in different parts of the sky was considered by Ambartsumyan (1936). The solution of this

problem using proper motions was obtained by Sobolev (1938). These methods, though of considerable theoretical significance, are not very suitable for practical applications. Let us consider a simpler, though less general technique, which gives the moments of the velocity distribution function from the observed radial velocities.

Let $d\omega$ be a small solid angle with galactic coordinates l and b, where l is reckoned from the direction to the center of the Galaxy. The number of stars with radial velocities between v_r and $v_r + dv_r$ in this solid angle is

$$N(l, b)\Phi(v_r, l, b)\, dv_r\, d\omega, \tag{35}$$

where $N(l, b)$ is the surface number density of stars, and $\varphi(v_r, l, b)$ is the normalized distribution function of radial velocities.

If u', v', and w' are the rectangular components of the star's velocity relative to the star, we have

$$v_r = \xi u' + \eta v' + \zeta w', \tag{36}$$

where ξ, η, and ζ are the direction cosines defined by

$$\xi = \cos l \cos b, \quad \eta = \sin l \cos b, \quad \zeta = \sin b.$$

Let further $-u_0, -v_0$, and $-w_0$ be the rectangular components of the residual velocity of the Sun, and u, v, and w the rectangular components of the star's residual velocity. Further suppose that we are dealing only with the nearest stars, so that the difference in the centroid velocities of the Sun and the stars is negligible and we may write

$$u = u' - u_0, \quad v = v' - v_0, \quad w = w' - w_0. \tag{37}$$

The average components of the residual velocities are by definition zero:

$$\bar{u} = 0, \quad \bar{v} = 0, \quad \bar{w} = 0, \tag{38}$$

and therefore

$$\bar{u}' = u_0, \quad \bar{v}' = v_0, \quad \bar{w}' = w_0. \tag{39}$$

Hence it follows that

$$\bar{v}_r = \xi u_0 + \eta v_0 + \zeta w_0. \tag{40}$$

If $F(u, v, w)$ is the sought distribution function of the residual velocity components, we have

$$N(l, b)\Phi(v_r, l, b)\, dv_r\, d\omega = N(l, b)\, d\omega \int_{(\Gamma)} F(u, v, w)\, d\gamma, \tag{41}$$

where $d\gamma = du'\, dv'\, dw'$ and the integral in the right-hand side of (41) is over the velocity in the velocity space between the two surfaces

$$\left.\begin{aligned} v_r &= \xi u' + \eta v' + \zeta w', \\ v_r + dv_r &= \xi u' + \eta v' + \zeta w'. \end{aligned}\right\} \tag{42}$$

Cancelling, we write (41) in the form

$$\Phi(v_r, l, b)\, dv_r = \int_{(\Gamma)} F(u, v, w)\, d\gamma. \tag{43}$$

Multiplying both sides of (43) by $(v_r - \bar{v}_r)^p$ we integrate over v_r from $-\infty$ to $+\infty$ remembering that relations (36) and (40) hold true for each v_r and $\bar{v}_r$, furthermore note that since (Γ) is the volume between two parallel planes distant dv_r from each other, integration in the right hand side over v_r from $-\infty$ to $+\infty$ in fact amounts to integration over the entire infinite space.

If we change over to the variables of integration u, v, and w, the volume of integration clearly remains the same. Thus,

$$\int_{-\infty}^{+\infty} (v_r - \bar{v}_r)^p \Phi(v_r, l, b)\, dv_r = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (\xi u + \eta v + \zeta w)^p F(u, v, w)\, du\, dv\, dw. \qquad (44)$$

For the central moment of p-th order of the radial velocity distribution function in the left-hand side of (44) we write $V_p(l, b)$. Let us further consider the corresponding moments of the function $\bar{F}$:

$$V_{i,j,k} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} u^i v^j w^k F(u, v, w)\, du\, dv\, dw. \qquad (45)$$

Seeing that

$$(a + b + c)^p = \sum_{i+j+k=p} \frac{p!}{i!j!k!} a^i b^j c^k \qquad (46)$$

and using in (44) the fundamental properties of integrals, we obtain

$$V_p(l, b) = \sum_{i+j+k=p} \frac{p!}{i!j!k!} \xi^i \eta^j \zeta^k V_{i,j,k}. \qquad (47)$$

Eq. (47) relates the moments of the distribution of radial velocities to the moments of residual velocities. The left-hand sides of Eqs. (47) are found from observation. The number of distinct equations for each p is clearly equal to the number of sky areas examined. Therefore, each of the equations in (47) yields a redundant system of equations, and the least squares method can be applied to find the moments of the residual velocity distribution function.

If F is symmetric, i.e., even in each of its arguments u, v, and w, then $V_{i,j,k}$ is zero if any of the subscripts i, j, k is odd.

Let us write out Eq. (47) for $p = 2$:

$$V_2(l, b) = \xi^2 V_{2,0,0} + \eta^2 V_{0,2,0} + \zeta^2 V_{0,0,2} + 2\xi\eta V_{1,1,0} + 2\xi\zeta V_{1,0,1} + 2\eta\zeta V_{0,1,1}. \qquad (48)$$

If the assumption of a symmetric distribution is valid, and the u, v, and w axes are the symmetry axes of the distribution, the coefficients $V_{1,1,0}$, $V_{1,0,1}$, and $V_{0,1,1}$ are so small that their sign will be unstable.

Thus if the distribution is indeed symmetric, terms with at least one odd subscript can be dropped in (47) for $p = 4$. Then

$$V_4(l, b) = \xi^4 V_{4,0,0} + \eta^4 V_{0,4,0} + \zeta^4 V_{0,0,4} + 6\xi^2\eta^2 V_{2,2,0} + 6\xi^2\zeta^2 V_{2,0,2} + 6\eta^2\zeta^2 V_{0,2,2}. \qquad (49)$$

If the residual velocities follow a Schwarzschild distribution, the moments obtained by the least squares method from (49) satisfy the approximate relations

$$\left.\begin{array}{lll} V_{2,2,0} = V_{2,0,0} \cdot V_{0,2,0}; & V_{2,0,2} = V_{2,0,0} \cdot V_{0,0,2}; & V_{0,2,2} = V_{0,2,0} \cdot V_{0,0,2}; \\ V_{4,0,0} = 3V_{2,0,0}^2; & V_{0,4,0} = 3V_{0,2,0}^2; & V_{0,0,4} = 3V_{0,0,2}^2. \end{array}\right\} \qquad (50)$$

The deviations of the actual moments from (50) provide an indication to what extent the Schwarzschild distribution is a good approximation to the actual residual velocity distribution.

Finally, Eq. (47) for $p = 3$,

$$V_3(l, b) = \xi^3 V_{3,0,0} + \eta^3 V_{0,3,0} + \zeta^3 V_{0,0,3}, \tag{51}$$

enables us to evaluate the skewness, if it is not particularly large. In this equality terms of the form $\xi^2\eta V_{2,1,0}$ and the term $\xi\eta\zeta V_{1,1,1}$ were dropped, since in the case of small skewness they are small compared to the retained terms. The skewness is assessed in terms of the ratios

$$\frac{V_{3,0,0}}{V_{2,0,0}^{3/2}}, \quad \frac{V_{0,3,0}}{V_{0,2,0}^{3/2}}, \quad \frac{V_{0,0,3}}{V_{0,0,2}^{3/2}}. \tag{52}$$

If the moments $V_{1,1,0}, V_{1,0,1}, V_{0,1,1}$ obtained from the solution of (48) are not small, this is either an indication of a skew distribution or, if the distribution is definitely symmetric (and the symmetry axes are orthogonal), this shows that the directions u, v, and w do not coincide with the symmetry axes of the distribution. Standard algebraic techniques should therefore be applied to transform the distribution to principal axes. To this end, we solve the so-called secular equation

$$\Delta\lambda = \begin{vmatrix} V_{2,0,0} - \lambda & V_{1,1,0} & V_{1,0,1} \\ V_{1,1,0} & V_{0,2,0} - \lambda & V_{0,1,1} \\ V_{1,0,1} & V_{0,1,1} & V_{0,0,2} - \lambda \end{vmatrix} = 0 . \tag{53}$$

The roots λ_1, λ_2, λ_3 are the reciprocal values of the dispersions of the velocity components along the principal axes, and the direction cosines l, m, n of these axes are obtained from the equations

$$\frac{l}{G_{11}} = \frac{m}{G_{12}} = \frac{n}{G_{13}} = \frac{1}{\sqrt{G_{11}^2 + G_{12}^2 + G_{13}^2}}, \tag{54}$$

where G_{11}, G_{12}, G_{13} are the minors of the elements of the first row of the determinant in (53) with λ_1, λ_2, λ_3 successively substituted for λ.

The moments of the residual velocity distribution can be similarly found using observations of proper motions or spatial velocities. These methods, however, are more complex and not as reliable as the methods based on radial velocities. If the proper motions are used, exact parallaxes of the stars are required. The parallaxes, however, are known only for few stars and even then with relatively large errors. The problem can be solved approximately if the parallaxes of the various stars are set equal to their common average parallax. This assumption, however, gives poor results.

When spatial velocities are used, it should be remembered that the starting observational material is unreliable, since the spatial velocity is found from observations of the radial velocity, two components of the proper motion and the parallax. All these quantities are usually determined with different relative errors, so that the available spatial velocities do not constitute a homogeneous sample.

§148. Asymmetry of stellar motions. The residual velocities of stars follow a symmetric hstribution if one takes into account the effect of observation errors and natural fluctuations. This property, however, is found to be characteristic only of speeds not exceeding 62 km/sec. Higher

residual speeds show a distinctly skewed distribution. For these speeds the allowed directions all lie in the range of galactic latitudes $130° < l < 340°$. The other directions covering about 40% of the sphere are "forbidden" for large velocities. It is significant that the skewness is not a continuous function of stellar velocities; in other words, it does not appear for moderate velocities gradually increasing toward the higher speeds, but it actually changes in a jump, from zero for velocities less than 62 km/sec to a finite and fairly uniform value for all velocities greater than 62 km/sec. This is clearly evident from Figure 170, borrowed from Oort's book (1926). The figure gives the galactic longitudes of the characteristic directions for three groups of velocities, $50 \leqslant v \leqslant 62$ km/sec, $63 \leqslant v \leqslant 99$ km/sec, $v \geqslant 100$ km/sec.

In all the diagrams, the radius of the dashed circle corresponds to 62 km/sec.

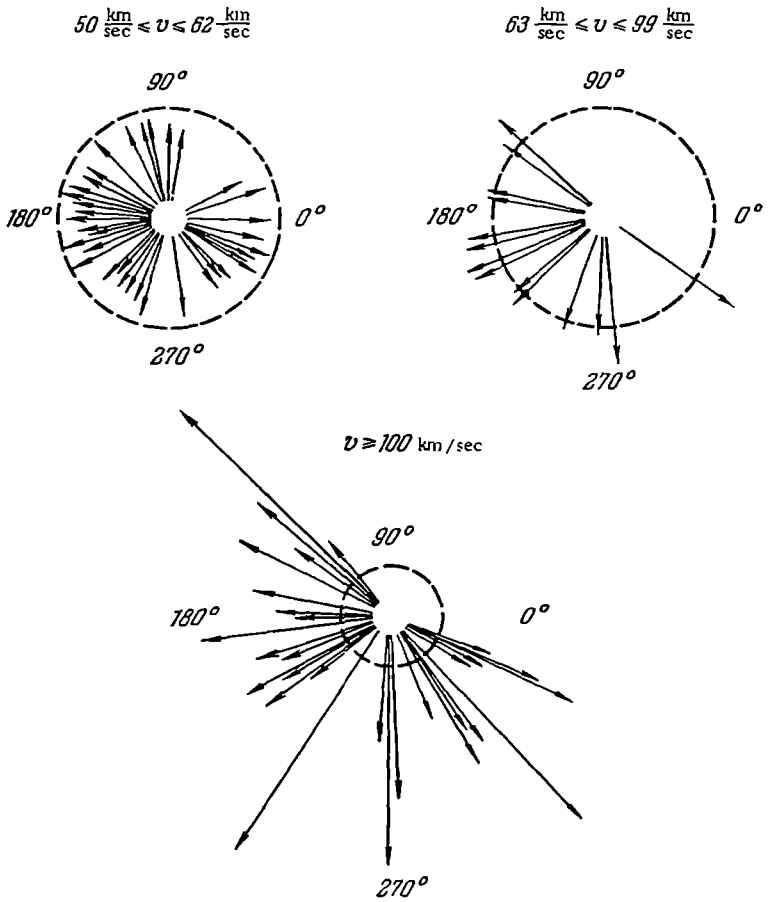


FIGURE 170

We see from the figure that velocities slightly less than 62 km/sec show an almost symmetric distribution, whereas velocities slightly above

62 km/sec have a distinctly skewed distribution; furthermore the skewness is nearly the same as for stars with $v \geqslant 100$ km/sec.

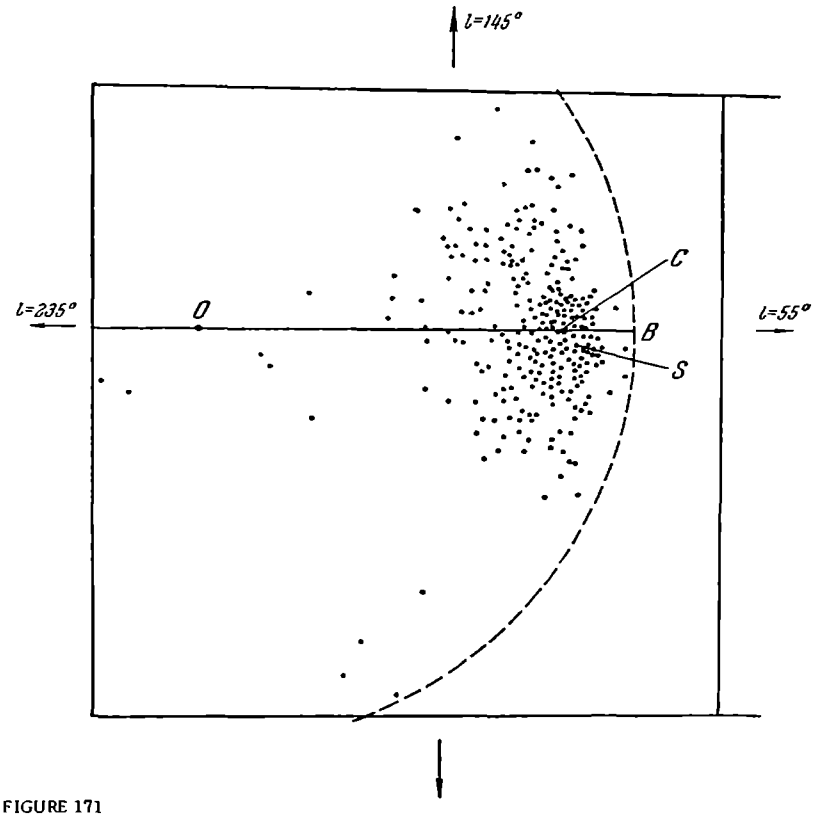


FIGURE 171

This phenomenon was first correctly interpreted by Lindblad /19/. He suggested that the Galaxy rotates around an axis perpendicular to the symmetry plane. The centroid of the Sun, in its motion around the center of the Galaxy, has a velocity directed to the point $l = 55°$, $b = 0°$. The speed of 62 km/sec is the difference between the critical speed in the neighborhood of the Sun (i.e., the speed sufficient for overcoming the gravitational attraction of the Galaxy and escaping into the intergalactic space) and the speed of the Sun's centroid. Therefore, if a residual velocity in excess of 62 km/sec were to show a direction of about $l = 55°$, the geometrical sum of this velocity and the centroid velocity would exceed the critical speed. A star with such a supercritical velocity would clearly leave the Galaxy. We see that even if stars with high residual velocities pointing in the sense of galactic rotation had existed at some time, they would have escaped from the Galaxy long ago, so that no such stars can be observed at present. High residual velocities pointing against the sense of galactic rotation are of course quite permissible, if we allow the rotation velocity of the Galaxy at the given point (i.e., the velocity of the Sun's centroid relative to the

center of inertia of the Galaxy) to be much greater than 62 km/sec. In Figure 171 each point represents the speed and the direction of a star in the galactic plane relative to the galactic center of inertia. Zero velocity is assigned to the point O. The velocity of the Sun's centroid is represented by the point C, having the direction $l = 55°$. The length of the radius-vector OC corresponds to the Sun's centroid speed $v_\odot$. Clearly, the radius-vectors which can be drawn from the point C to the various points in the diagram determine the magnitude and the direction of the residual velocities of the corresponding stars. Thus, the radius-vector CS, 19.5 km long and directed to the point with $l = 23°.5$, corresponds to the individual velocity of the Sun. The radius of the dashed circle around the point O in the figure is equal to the critical velocity v_p. The segment CB corresponds to 62 km/sec. The distribution of the points in the diagram is such that none of them lie outside the critical velocity circle. This clearly requires an asymmetric placement of the distant points relative to the point C, whereas the nearby points show a symmetric distribution. Figure 171 reveals that stars with large residual velocities move comparatively slowly relative to the galactic center of inertia. Finally note that it is because v_p is much greater than 62 km/sec ($OB \gg CB$) that the skewness shows abruptly at residual velocities above 62 km/sec and hardly increases at higher velocities. It is readily seen that if, for instance, $\frac{OB}{CB}$ were equal to 2, the skewness would follow a different pattern.

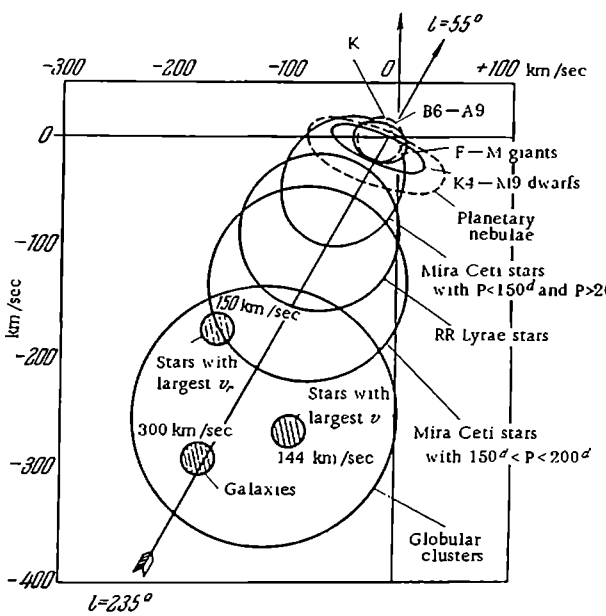


FIGURE 172

Stromberg (1925) compared the centroid velocities of various groups of objects (special centroids) with their dispersions. The results are shown in Figure 172. Here the center of each circle corresponds to the centroid

velocity, and the radius of the circle is equal to the dispersion of the velocities of the corresponding objects. The centroid velocities are taken relative to the Sun, and not the Sun's centroid. The actual radii of the cross-hatched circles are so large that they cannot be shown to scale in the diagram; we therefore represent them by small circles, indicating the corresponding radius. The following two regular features are clearly observed: 1) the centers of all the circles fall along a single straight line; 2) the greater the centroid velocity relative to the Sun, the higher is the dispersion of the velocities of the corresponding objects.

These features can be interpreted as follows. Let v_c be the centroid velocity relative to the center of inertia of the stellar system, v the residual velocity of the object, w the total velocity, i.e., the velocity relative to the center of inertia. Then by definition

$$w = v_c + v. \tag{55}$$

Squaring and averaging we obtain

$$\overline{w^2} = v_c^2 + 2\overline{v_c \cdot v} + \overline{v^2}. \tag{56}$$

If the residual velocities are symmetrically distributed, the scalar averages out ot zero, and we get

$$\overline{w^2} = v_c^2 + \overline{v^2}, \tag{57}$$

where v^2 is the dispersion of the residual velocities.

Now suppose that the mean square of the total velocities is the same for the different subsystems. Then the maximum centroid velocity corresponds to the subsystem with zero dispersion. This centroid velocity v_{c_0} is represented by the point K in Figure 172. If now the directions of all the special centroids coincide, we see from (57) that (a) the higher the disperison of velocities in the subsystem the greater is the difference between v_{c_0} and the centroid velocity and (b) all the vector differences between the velocities v_{c_0} and the subsystem centroids fall along one line.

The skewness of the large velocities is also responsible for the pattern observed in Figure 171. Indeed, if equal residual speeds are considered, we see from Figure 171 that the higher the speed, the greater the dispersion of the residual velocities. The figure shows, however, that the greater the speed the less is the mean value of the projections of the corresponding total velocities on the OC axis, i.e., the less is the centroid velocity.

It follows from the preceding that if we assume a rotating Galaxy, the direction of the line OC in Figure 171 and of the straight line in Figure 172 should coincide with the direction of galactic rotation.

The asymmetry of stellar motions thus provided the first indication of galactic rotation.

§149. The effect of galactic rotation on radial velocities and proper motions. The field of force responsible for the motion of stars in the Galaxy is internally generated, i.e., the Galaxy is a self-gravitating system. The law of galactic rotation is therefore a function of the distribution of matter density in the Galaxy.

If our stellar system had a uniform density, it would rotate almost as a rigid body, with all the centroids revloving about the center of the Galaxy with nearly constant angular velocity. In this case, the distance between

any two centroids would remain constant, and the galactic rotation would have no effect on the radial velocities of stars; it would not affect the proper motions of stars either. This is not so, however. Clearly, if the Galaxy rotated as a rigid body about an axis through the galactic pole, the point on the celestial sphere defined by the direction from one centroid to another would move uniformly along the galactic parallel with a period equal to the galactic rotation period. Since the galactic system of coordinates, like the equatorial system, is fixed, the rigid-body rotation of the Galaxy would add a term equal to ω (the angular velocity of ratation of the Galaxy in seconds per year) to the proper motion in galactic longitude μ_l. This term would be the same for all the stars and it could not be detected in differential measurements of proper motions, although it would immediately emerge from absolute measurements of proper motions. The latitudinal proper motion μ_b would not be affected by the rigid-body rotation of the Galaxy.

If the entire mass of the Galaxy were concentrated at its center, the stars would revolve in Keplerian orbits. This would also approximately hold true for the Sun and the nearest stars, provided that they were located outside the bulk of the Galaxy. In Keplerian motion, as we know, the mean angular velocity of revolution rapidly decreases with distance from the center.

In other star systems, galaxies and star clusters, the stellar density or, equivalently, the surface luminosity (in yellow and red light) on the whole decrease away from the center. By inference this property is also applicable to our Galaxy, so that in fact we are probably dealing with an intermediate case between the two extremes of a homogeneous density system and a central mass concentration. This signifies that the rotation velocity in the Galaxy should decrease away from the center, although not as fast as prescribed by Kepler's law.

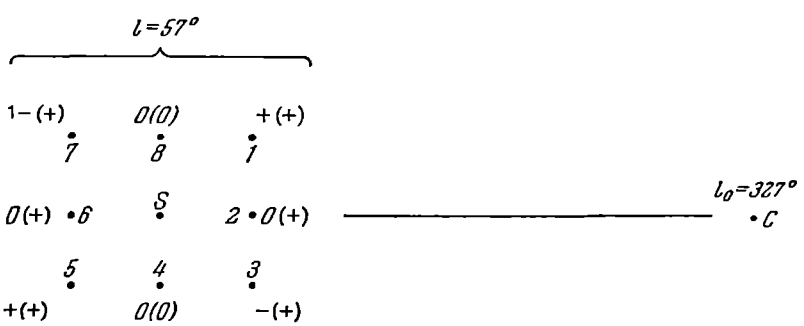


FIGURE 173

Figure 173 presents eight numbered centroids which are located symmetrically relative to the Sun's centroid S in the plane of the Galaxy. The galactic center is to the right, in the direction $l = 327°$. The galactic rotation is at right angles to the direction to the center, $l = 57°$, say. The centroids 1, 2, 3, closest to the center of the Galaxy, should have higher angular velocities than the centroids 4, S, 8, and these three should have higher angular velocities than the centroids 5, 6, 7. In this case, as is easily seen centroid 1 will be carried progressively farther away from S

by the differential rotation of the Galaxy, i.e., it will have a positive radial velocity; the instantaneous distance of centroid 2 from S does not change, i.e., its radial velocity is zero, etc. If the radial velocity of the centroid is not zero, its sign is marked outside the parentheses in the figure; otherwise a nought is written. We see that the radial velocity caused by the differential rotation is characterized by a double longitudinal wave of the form $\sim\sin(l - l_0)$.

Since the centroids 1, 2, 3 catch up with S and the centroids 5, 6, 7 lag behind, the differential rotation of the Galaxy should clearly introduce a positive term to the longitudinal proper motion of all these centroids; the largest positive contribution is for centroids 2 and 6, whereas for centroids 4 and 8 this contribution is zero. The sign of the contribution is shown in parentheses in the figure. The proper motion produced by the differential rotation of the Galaxy is also described by a double longitudinal wave, though of a different form, $\sim 1 + \cos(l - l_0)$.

The differential rotation of the Galaxy can be interpreted as the rotation which is obtained on subtracting the observer's angular velocity from the angular velocity of each centroid. The galactic rotation is therefore made up of rigid-body rotation with observer's centroid angular velocity plus differential rotation. We have so far only considered the effect of differential rotation. On passing to the rigid-body component, we should remember that it does not affect the radial velocities, whereas the longitudinal proper motion introduces a constant term for all the centroids, equal in magnitude to ω. According to Figure 173, this term is negative. Thus, if rigid-body rotation is taken into consideration, the parenthetical sign for centroids 4 and 8 in Figure 173 should be minus, and for the rest of the centroids it depends on the relative magnitude of the two terms: the positive one due to the differential rotation and the negative one produced by the rigid-body rotation of the Galaxy.

§150. The differential field of centroid velocities. Its effect on the observed stellar motions. In the previous section we have qualitatively considered the terms in radial velocities and proper motions of stars assuming that all the centroids in the Galaxy simply revolve around its center. A quantitative solution of this problem was first proposed by Oort (1928), who can thus be regarded as the initiator of the study of galactic rotation.

A more general problem was considered in 1932 by Ogorodnikov, who assumed arbitrary motions of centroids in planes parallel to the plane of the Galaxy. In 1935 Milne generalized Ogorodnikov's theory to the case of centroid motion in three dimensions.

Let $\Delta v\,(x_1, x_2, x_3)$ be the vector difference in the centroid velocities between the field point (x_1, x_2, x_3) and the observer at $(0, 0, 0)$. The x_1 axis is directed to the center of the Galaxy, the x_2 axis at right angles to x_1 in the plane of the Galaxy ($l = 55°$), and x_3 parallel to the galactic symmetry axis. Let u_1, u_2, u_3 be the components of the vector Δv in this frame.

Further consider a moving frame (r, l, b), where the first axis is directed to the point (x_1, x_2, x_3), the second axis lies in the plane of the sky in the direction of increasing galactic longitudes, and the third axis in the plane of the sky in the direction of increasing galactic latitudes. The projections of Δv on these axes respectively give Δv_r — the radial velocity of the centroid,

Δv_l — the projection of its transversal velocity in the direction of increasing galactic longitudes, and Δv_b — the projection of the transversal velocity in the direction of increasing latitudes. These projections are expressed by the relations

$$\Delta v_r = \sum_{i=1}^{3} \lambda_{r,i} u_i, \qquad \Delta v_l = \sum_{i=1}^{3} \lambda_{l,i} u_i, \qquad \Delta v_b = \sum_{i=1}^{3} \lambda_{b,i} u_i. \tag{58}$$

where the directions cosines λ are arranged in Table 5.

TABLE 5

i	1	2	3
$\lambda_{r,i}$	$\cos b \cos(l-l_0)$	$\cos b \sin(l-l_0)$	$\sin b$
$\lambda_{l,i}$	$-\sin(l-l_0)$	$\cos(l-l_0)$	0
$\lambda_{b,i}$	$-\sin b \cos(l-l_0)$	$-\sin b \sin(l-l_0)$	$\cos b$

In the table, l and b are the galactic coordinates of the direction to (x_1, x_2, x_3), and l_0 is the longitude of the galactic center.

Now, since the centroid field is continuous and $r = \sqrt{\sum_{i=1}^{3} x_i^2}$ can be taken sufficiently small, we expand $u_i(x_1, x_2, x_3)$ in a Taylor series and retain only the first-order terms:

$$u_i = \sum_{j=1}^{3} x_j \left(\frac{\partial u_i}{\partial x_j}\right)_0. \tag{59}$$

Since

$$\frac{x_j}{r} = \lambda_{r,j},$$

Eq. (59) can be written in the form

$$u_i = r \sum_{j=1}^{3} \lambda_{r,j} \left(\frac{\partial u_i}{\partial x_j}\right)_0. \tag{60}$$

Inserting (60) in (58), we get

$$\left.\begin{aligned} \Delta v_r &= r \sum_{j=1}^{3} \sum_{i=1}^{3} \lambda_{r,i} \lambda_{r,j} \left(\frac{\partial u_i}{\partial x_j}\right)_0, \\ \Delta v_l &= r \sum_{j=1}^{3} \sum_{i=1}^{3} \lambda_{l,i} \lambda_{r,j} \left(\frac{\partial u_i}{\partial x_j}\right)_0, \\ \Delta v_b &= r \sum_{j=1}^{3} \sum_{i=1}^{3} \lambda_{b,i} \lambda_{r,j} \left(\frac{\partial u_i}{\partial x_j}\right)_0. \end{aligned}\right\} \tag{61}$$

Eqs. (61) solve our problem. The unknowns here are the nine quantities $\left(\frac{du_i}{dx_j}\right)_0$ which characterize the centroid field in the neighborhood of the Sun. If for each equation in (61) we consider more than nine sky areas and in each area determine the average radial velocity of the stars and the average proper motions in galactic latitude and galactic longitude μ_b and μ_l, we obtain three sets of redundant equations. The second and the third equations in (61) are modified, seeing that

$$\frac{\Delta v_l}{r} = 4.738\,\mu_l \cdot \sec b,$$

$$\frac{\Delta v_b}{r} = 4.738\,\mu_b.$$

The coefficients λ are calculated for the area centers. In solving the redundant equations (61) the residual velocities are regarded as random errors and therefore affect only the mean error of the results.

The first equation requires knowledge of the distance r. Nevertheless, it is the first equation that gives the most reliable results, since the average proper motions are greatly distorted by individual nearby stars with large proper motions, and the relative errors in small proper motions are large.

Following Ogorodnikov, let us consider in more detail a particular case of plane-parallel motion. Then $u_3 = 0$. Moreover, if we regard the Sun as located in the plane of the Galaxy, we may write

$$\left(\frac{\partial u_1}{\partial x_3}\right)_0 = \left(\frac{\partial u_2}{\partial x_3}\right)_0 = 0,$$

and Eqs. (61) in expanded form may be written

$$\left.\begin{aligned} \Delta v_r &= r\cos^2 b\,[K + C\cos 2(l-l_0) + A\sin 2(l-l_0)], \\ \Delta v_l &= r\cos b\,[B + A\cos 2(l-l_0) + C\sin 2(l-l_0)], \\ \Delta v_b &= -r\sin b\cos b\,[K + A\cos 2(l-l_0) + C\sin 2(l-l_0)], \end{aligned}\right\} \qquad (62)$$

where

$$\left.\begin{aligned} A &= \frac{1}{2}\left[\left(\frac{\partial u_1}{\partial x_2}\right)_0 + \left(\frac{\partial u_2}{\partial x_1}\right)_0\right]; & B &= \frac{1}{2}\left[\left(\frac{\partial u_2}{\partial x_1}\right)_0 - \left(\frac{\partial u_1}{\partial x_2}\right)_0\right], \\ K &= \frac{1}{2}\left[\left(\frac{\partial u_1}{\partial x_1}\right)_0 + \left(\frac{\partial u_2}{\partial x_2}\right)_0\right]; & C &= \frac{1}{2}\left[\left(\frac{\partial u_1}{\partial x_1}\right)_0 - \left(\frac{\partial u_2}{\partial x_2}\right)_0\right]. \end{aligned}\right\} \qquad (63)$$

We see that $2K$ is div $\Delta\vec{v}$, and $2B$ is equal to rot $\Delta\vec{v}$. Thus, K characterizes expansion or contraction of a small volume around the observer (depending on its sign), and B describes the rotation of this volume about the observer.

As for the constants A and C, writing $A = N\cos 2\psi$, $C = N\sin 2\psi$, we may represent the sum of the second and third term in Δv_r in the form $N\sin 2(l - l_0 + \psi)$ and that in Δv_l in the form $N\cos 2(l - l_0 + \psi)$. Thus, the term C is equivalent to a constant correction to l_0. We can thus omit the term C, treating l_0 as an unknown. Then we readily see that because of the term A, points originally lying on the circle $x_1^2 + x_2^2 = r^2$ will occupy after unit time a curve which is described in polar coordinates by

$$\varrho = r\,[1 + A\sin 2(l - l_0)]. \qquad (64)$$

Eq. (64) is the equation of a closed curve stretched in the direction $l_0 \to l_0 + 180°$ if $A > 0$ and compressed in the same direction for $A < 0$. The coefficient A thus characterizes the deformation of a small volume about the observer.

Application of Eqs. (62) to the nearest stars showed that the coefficient K was virtually zero for all stars, except those of spectral type B, for which we have $K = 4.2$ km/sec (for stars brighter than 6^m). For weaker B stars, K is much less.

Thus, if the bright B stars constitute but a minor fraction in a given statistical ensemble, Eqs. (62) may be written in the form

$$\left.\begin{aligned} \Delta v_r &= rA \sin 2(l - l_0)\cos^2 b, \\ \Delta v_l &= rA \cos 2(l - l_0)\cos b + rB\cos b, \\ \Delta v_b &= -rA \sin 2(l - l_0)\sin b \cos b. \end{aligned}\right\} \tag{65}$$

We see that the coefficient B cannot be determined from the radial velocities and from the proper motion component μ_b.

Now suppose that the centroid field corresponds to pure rotation of the Galaxy about its axis, i.e., each centroid moves in a circular orbit parallel to the galactic plane with angular velocity which is only a function of the distance from the rotation axis, $\omega(R)$.

The coefficients A and B entering Eq. (65), which are called the Oort coefficients, can be expressed in terms of ω and R. In circular motion $u_2 =$. Since $R = R_0 - x_1$, we have

$$\left(\frac{\partial u_2}{\partial x_1}\right)_0 = -\left[\frac{d(\omega R)}{dR}\right]_0.$$

To find $\left(\frac{\partial u_1}{\partial x_2}\right)_0$ note that after a displacement Δx_2, the centroid velocity is no longer perpendicular to the x_1 axis and the projection of the centroid velocity on this axis is $+\omega R \sin\frac{\Delta x_2}{R}$. Therefore

$$\left(\frac{\partial u_1}{\partial x_2}\right)_0 = \left[\lim_{\Delta x_2 \to 0} \frac{\omega R \sin\frac{\Delta x_2}{R}}{\Delta x_2}\right] = +\omega_0.$$

Thus

$$\left.\begin{aligned} A &= \frac{1}{2}\left[-\frac{d(\omega R)}{dR} + \omega\right]_0 = -\frac{1}{2}\left[R\frac{d\omega}{dR}\right]_0, \\ B &= \frac{1}{2}\left[-\frac{d(\omega R)}{dR} - \omega\right]_0 = A - \omega_0. \end{aligned}\right\} \tag{66}$$

Using the equality $B = A - \omega_0$ and the relations between tangential velocities and proper motions, we write (65) in the form

$$\begin{aligned} \Delta v_r &= rA \sin 2(l - l_0)\cos^2 b, \\ \Delta\mu_l &= \frac{1}{4.74} A\,[1 + \cos(l - l_0)] - \frac{1}{4.74}\omega_0, \\ \Delta\mu_b &= -\frac{1}{4.74} A \sin 2(l - l_0)\sin b \cos b. \end{aligned}$$

We see that the first two of these equalities correspond to the expressions predicted by the qualitative treatment in §145.

The Oort coefficients have the dimension of angular velocity (sec $^{-1}$). They are generally expressed in km/sec per parsec. The determination of these coefficients and of the constant l_0 from (65) was the subject of numerous studies.

Note that although K is equal to zero and A and B are finite, this does not constitute a proof that the centroid field of the Galaxy is a field of pure rotation about the center of symmetry. This only proves that the small volume around the Sun neither expands nor contracts, although it is subjected to distortion and rotation. On the other hand, the fact that $C = 0$ and l_0 in (65) coincides with the galactic longitude of the center of the Galaxy as determined by alternative methods provides a substantial argument in favor of galactic rotation. This shows that the rotation of the volume near the Sun is associated with the revolution of that volume around the center of the Galaxy. Another weighty argument in favor of galactic rotation is provided by the good fit between the value of ω_0 derived from (66),

$$\omega_0 = A - B,$$

and the definition of angular velocity

$$\omega_0 = \frac{v_0}{R_0},$$

where v_0 is the Sun's centroid velocity inferred from extragalactic objects, and R_0 is the distance to the center of the Galaxy, determined, e.g., as the distance to the center of the system of globular clusters.

Table 6 lists various determinations of A, B, and l_0 (Parenago's and more recent findings). Here A and B are given in km/sec per kpc.

TABLE 6

Author	Year	Sample	n	A	B	l_0
Raymond, Wilson	1931	v (all types)	4233	+15	−16	329°
Plaskett, Pearce	1934	v_r (O—B7)	849	+21	—	324
Plaskett, Pearce	1934	μ (O—B7)	717	+13	−12 ± 3	312
Nordstrom	1936	v_r (A—M)	847	+21 ± 5	—	336 ± 7
Oort	1937	μ of faint stars	18,000	—	−15 ± 2	—
Van de Kamp, Vysotskii	1937	μ of faint stars	18,000	+14 ± 3	−14 ± 3	321
Wilson, Raymond	1938	μ from Boss' catalogue	33,432	+12	−12	326
Joy	1936	v_r (cepheids)	156	+21	—	325
Gliese	1940	μ from FK3 catalogue	1535	+13 ± 6	−912	317 ± 12
Parenago	1941	v_r (cepheids, O—B, and others)	769	+19.7 ± 4	—	—
Vysotskii, Williams	1948	μ from FK3 catalogue	69,000	+19.0 ± 1.4	−7.6 ± 1.4 −9	— 330
Shatsova	1950	GC	31,922	+18		
Morgan, Oort	1951	FK3, N30		+20	−7	
Bakulin	1953	Supergiants and long-period cepheids, GC	870	+14 ± 1	−12 ± 1 −8 ± 1	317 ± 3 320 ± 3
Bakulin	1953	Same stars of the FK3 catalogue	870	+15 ± 1	—	—
Weaver	1955	v_r (cepheids)	79	+10.8 ± 1.8	—	—
Weaver	1955	v_r (O9.5—B2)	187	+13.2 ± 2.7	—	—
Petrie, Cuttle, Andrews	1956	v_r (B stars and others)	143	+17.7 ± 1.1	—	—
Stibbs	1956	v_r (cepheids)	55	+19.5 ± 1.9	—	—
Gascoigne, Eggen	1957	v_r (cepheids)	55	+17.5 ± 1.9	—	—

Table 6 shows that A, B and l_0 determined from various objects are surprisingly stable. The probable values of Oort's coefficients and l_0 are thus

$$\left.\begin{aligned} A &= +0.018 \text{ km/sec/pc} = +0''.0038 \text{ annually}, \\ B &= -0.012 \text{ km/sec/pc} = -0''.0025 \text{ annually}, \end{aligned}\right\} \qquad (67)$$
$$l_0 = 325^\circ.$$

The angular velocity of rotation is found to be

$$\omega_0 = A - B = 0''.0063 \text{ annually} = 9.7 \cdot 10^{-16} \text{ rad/sec}, \qquad (68)$$

which corresponds to a galactic rotation period of 205 million years. The average velocity of the Sun relative to extragalactic objects, corrected for the residual velocity, is found to be $v_0 = 220$ km/sec. Its direction is almost precisely perpendicular to the direction to the galactic center. This therefore seems to be the velocity with which the Sun's centroid revolves around the galactic center, and the possible residual velocity of the Galaxy can be ignored. In this case we obtain for the distance to the center of the Galaxy

$$R_0 = \frac{v_0}{\omega_0} = \frac{220 \frac{\text{km}}{\text{sec}}}{9.7 \cdot 10^{-18} \text{ sec}^{-1}} = 2.3 \cdot 10^{17} \text{ km} = 7400 \text{ pc}.$$

This result is in good agreement with alternative determinations of the distance to the galactic center.

What information do the coefficients A and B provide on the character of galactic rotation in the Sun's neighborhood? If the density of our star system were uniform, the Galaxy would rotate as a rigid body. In this case, assuming the Sun to be located inside the system, we would have

$$\omega(R) = \text{const}; \quad \omega'(R) = 0; \quad A = 0; \quad B = -\omega_0. \qquad (69)$$

If the Sun were located outside the stellar system it would move in a Keplerian orbit around the center of the Galaxy together with other stars in its neighborhood:

$$\left.\begin{aligned} &\omega \sim R^{-\frac{3}{2}}; \quad \omega' = -\frac{3}{2}\frac{\omega}{R}; \\ &A = \frac{3}{4}\omega_0; \quad B = -\frac{1}{4}\omega_0. \end{aligned}\right\} \qquad (70)$$

The results (67) and (68) based on observations in fact give

$$A = 0.6\,\omega_0; \quad B = -0.4\,\omega_0. \qquad (71)$$

Comparison of (71) with (69) and (70) shows that the motion of our stellar system in the Sun's neighborhood is intermediate between rigid-body rotation and Keplerian revolution, approaching the latter.

We thus conclude that the Sun is situated inside the Galaxy but at its periphery, so that the stellar density in regions closer to the center of the Galaxy is substantially higher than that outside the orbit of the Sun's centroid.

§151. The centroid field for pure galactic rotation. In the previous section, the centroid field was investigated using the proper motions of the nearest stars. Therefore, besides the angular velocity of the Sun's centroid, the results only contain the first derivative of this velocity.

The results of §145 seem to indicate that the centroid field is a field of pure rotation around the axis of symmetry of the Galaxy. Proceeding from this assumption, we can assess its influence on the motion of stars at any distance.

The velocity vector of a centroid S relative to the Sun's centroid S_0 is given by the difference of vector products

$$\Delta v = [\omega \times R] - [\omega_0 \times R_0], \tag{72}$$

where nought identifies the quantities corresponding to the Sun's centroid.

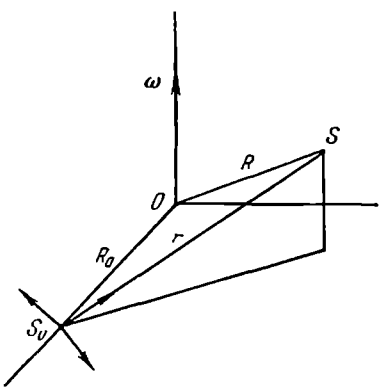


FIGURE 174

Introducing a topocentric radius-vector r of the centroid S relative to the centroid S_0 (see Figure 174), we may write

$$R = R_0 + r. \tag{73}$$

Inserting (73) in (72), we obtain

$$\Delta v = [\omega - \omega_0 \times R_0] + [\omega \times r]. \tag{74}$$

The magnitude of the first vector in the right-hand side of (74) is $(\omega - \omega_0)R_0$ and it points in the direction of galactic rotation. Its direction cosines in the moving frame (r, l, b) are described by the expressions listed in the second column of Table 5. The magnitude of the second vector in the right-hand side of (74) is $\omega r \cos b$ and it points in the direction of increasing longitudes, i.e., along the l axis in the (r, l, b) frame.

By the preceding, the projections of the vector (74) on the (r, l, b) axes are

$$\left.\begin{aligned} \Delta v_r &= (\omega - \omega_0) R_0 \cos b \sin (l - l_0), \\ \Delta v_l &= (\omega - \omega_0) R_0 \cos (l - l_0) + \omega r \cos b, \\ \Delta v_b &= -(\omega - \omega_0) R_0 \sin b \sin (l - l_0). \end{aligned}\right\} \tag{75}$$

Eqs. (75) are in principle suitable for studying the motion of distant centroids. However, since the proper motions of the far objects are determined with a high degree of uncertainty, only the first of these equations can actually be applied to distant objects. It was first used by Camm (1938). From the first equation in (75) and Figure 174 we get

$$\left.\begin{aligned} &\omega(R) - \omega(R_0) = \Delta v_r \cdot \sec b \cdot \operatorname{cosec}(l - l_0), \\ &R = [R_0^2 + r^2 \cos^2 b - 2R_0 r \cos b \sin(l - l_0)]^{\frac{1}{2}}. \end{aligned}\right\} \qquad (76)$$

Thus, assuming certain numerical values for R_0 and l_0 and measuring Δv_r and r, we can find $\omega(R) - \omega(R_0)$ as a function of R.

§152. Galactic rotation from radio observations. Some data on the distribution of gaseous matter in the Galaxy from radio observations were described in §134. Observations of neutral hydrogen at 21 cm wavelength (a unique emission line of HI) also provide valuable information on the motion of neutral hydrogen clouds in different parts of the Galaxy.

Note that neutral hydrogen moves in the same gravitational field in which all the other objects of our stellar system move. Therefore, information on its motion is of great importance for the study of galactic dynamics on the whole. On the other hand, the motion of neutral hydrogen may have its specific features too.

In 1954 van de Hulst, Muller and Oort published their measurements of the 21-cm line profiles for 54 directions in the plane of the Galaxy between galactic longitudes of 322 and 220° using a 7.5-m radio telescope. In the same year, Kwee, Muller, and Westerhout repeated these measurements for a region between 321°.1 and 45°.0 galactic longitude using the same instrument with improved equipment, so that more detailed profiles were obtained. Figure 175 shows the profiles of 30 lines obtained by these authors at longitudes between 321°.1 and 33°.4. The measurements were carried out at latitudes of maximum intensity at each longitude. The corresponding latitudes were −1°.4, −1°.5, −1°.6, or −1°.7 (in the old system of galactic coordinates) for the different longitudes. These latitudes and the respective longitudes are marked in Figure 175. The horizontal axis gives the radial velocities v_r in km/sec (corrected for the Earth's velocity relative to the centroid of the nearest stars), which can be converted to wavelengths using Doppler's equation. The vertical axis gives the radiation intensity I which can be expressed on the absolute temperature scale.

By analogy with stellar motions, it seems that the particles of neutral hydrogen have definite residual velocities, in addition to the circular motion of the centroid about the center of the Galaxy. This assumption is clearly confirmed by the shape of the centerward line profile. In case of pure circular motion, the hydrogen particles which lie in the direction to the center of the Galaxy would have $v_r = 0$ relative to the Sun's centroid, and this would produce a sharp peak; the line width would be virtually zero. In fact, the lines observed in directions close to centerward have significant widths with effective values of about 40 km/sec. A similar situation, according to van de Hulst, Muller, and Oort, is observed in the direction to the anticenter of the Galaxy, where in case of pure circular motion we should expect almost monochromatic radiation corresponding to $v_r = 0$.

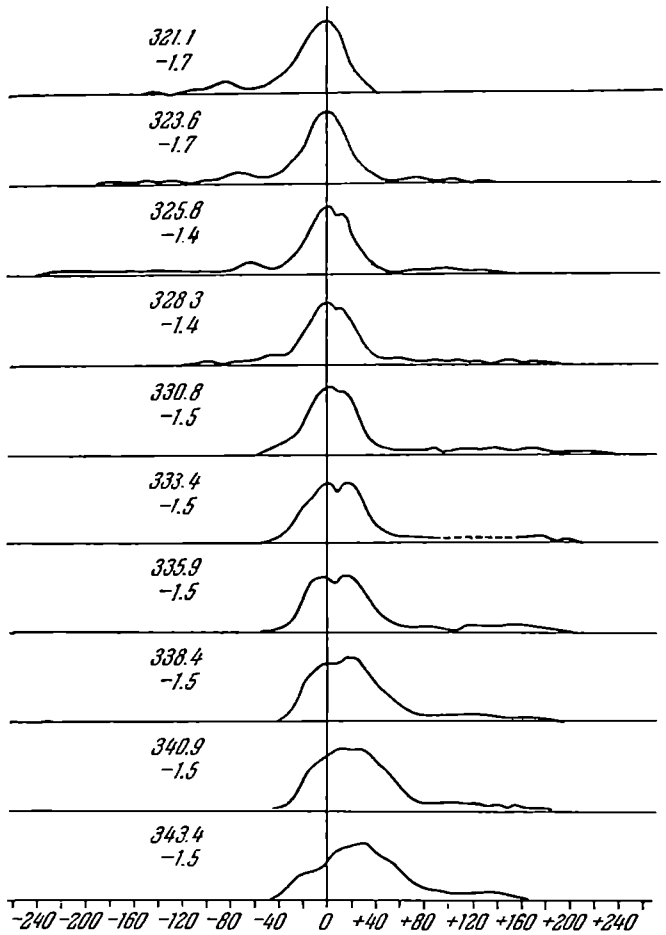


FIGURE 175a

Although this proves conclusively that neutral hydrogen particles have definite residual velocities, we will first consider pure circular motion of the gas. Let the point S in Figure 176 correspond to the Sun, the point C to the center of the Galaxy. The point P lies in the observed direction in the plane of the Galaxy (characterized by the angle $l' = l - l_0$ at a distance r from S. Let ω_0 and $\omega(R)$ be the angular velocity of rotation of the Sun's centroid and the centroid of the point P, respectively. The radial velocity of P is then given by

$$v_r = R\omega(R)\sin\varphi - R_0\omega_0 \sin l'. \tag{77}$$

Inserting for sin φ from the theorem of the sines, we obtain

$$v_r = R_0[\omega(R) - \omega_0]\sin l'. \tag{78}$$

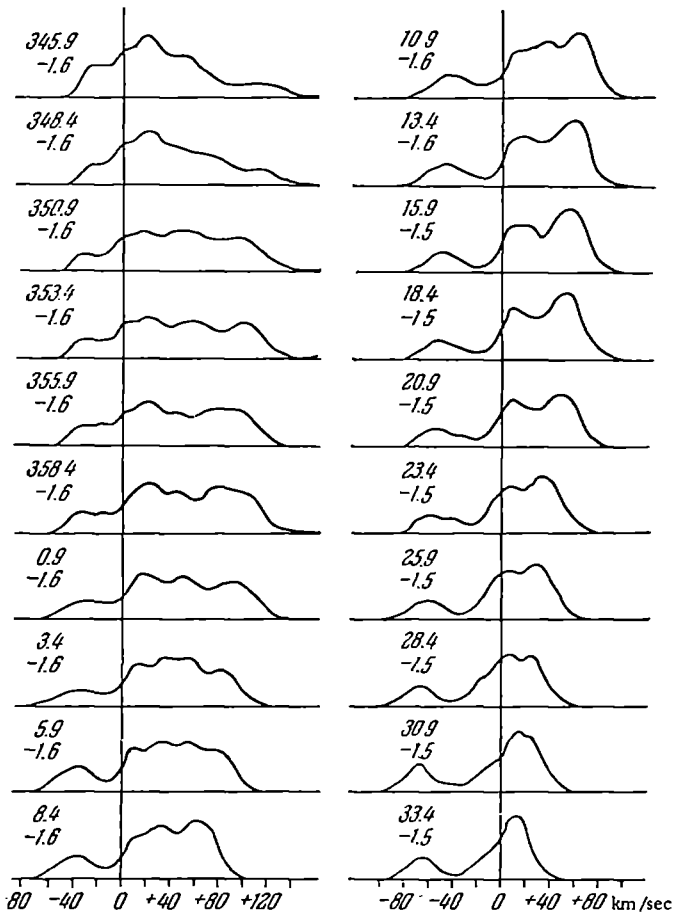


FIGURE 175b

Galactic rotation data from observations of stars, the rotation of other galaxies, and theoretical considerations indicate that the angular velocity of the centroids monotonically decreases with distance from the center of the Galaxy. Therefore in Figure 176 all the points between S and K should have positive radial velocity by (78), reaching a maximum at the point D, the base of the perpendicular. Points lying outside the segment SK, on the other hand, should have negative radial velocity. In Figure 176, $0 < l' < 90°$. For $90° < l' < 180°$, as is easily seen, the velocities of all the points along the line of sight are negative, and for $180° < l' < 270°$ they are positive. If $270° < l' < 360°$, the sign of the radial velocity is the opposite of that for $0 < l' < 90°$.

If the neutral hydrogen were uniformly distributed in the plane of the Galaxy, we could assume circular motion and ignore the absorption at the wavelength of its emission lines to derive a simple theoretical expression for the line profile in terms of the known function $\omega(R)$ and the distance of

the Sun from the center of the Galaxy R_0. Indeed, in a given direction, according to Eq. (78), the number of hydrogen atoms in any range of v_r values increases as the variation of $\omega(R)$ with r in the given locality becomes slower. In other words,

$$I = \frac{g}{\frac{d\omega}{dr}}, \tag{79}$$

where g is proportional to the density of hydrogen and inversely proportional to $R_0 \sin l'$. But

$$\frac{d\omega}{dr} = \frac{d\omega}{dR}\frac{dR}{dr} = \omega'(R) \cdot \sqrt{1 - \frac{R_0^2}{R^2}\sin^2 l'},$$

so that

$$I = \frac{g}{\omega'(R) \cdot \sqrt{1 - \frac{R_0^2}{R^2}\sin^2 l'}} \tag{80}$$

and the dependence of I on v_r is expressed in terms of the parameter R by (78) and (80).

If $0 < l' < 90°$, as is the case for most of the line profiles in Figure 175, the maximum positive v_r is attained at D. At the same point, I is infinitely large by (80), since here $R = R_0 \sin l'$. As we move away from P in either direction, v_r drops to zero inside the segment SK; I also decreases, since the radicand in the denominator in (80) increases and $\omega'(R)$ is of bounded variation. Beyond K, v_r is negative, and I is immediately halved, since inside the segment SK there are two points corresponding to each v_r, whereas beyond K there is but one point corresponding to each v_r. As v_r further decreases (in relative terms), the intensity should decrease.

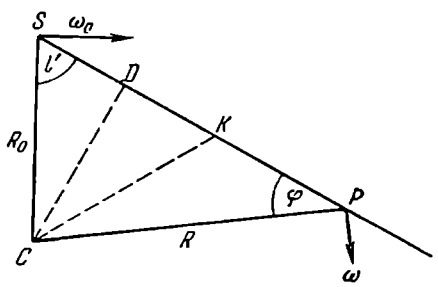


FIGURE 176

The observed lines (Figure 175) behave on the whole as prescribed by Eqs. (78) and (80). Note, for instance, that for $20° < l' < 50°$ the overall line intensity in the region $v_r > 0$ is much greater than it is for $v_r < 0$, and for $l' > 50°$ this relation is gradually reversed in favor of the region $v_r < 0$; this is associated with the fact that the distance SK (corresponding to points with $v_r > 0$) is large for the first group of l' values and small for $l' > 50°$.

However, despite the overall fit between Eqs. (78) and (80) and the observed profiles, significant deviations from these formulas are observed. First, we see no high intensity peak for the maximum v_r with the I falling off steeply to zero to the right of this v_r. The acutal line is somewhat spread. apparently due to the effect of the residual velocities of neutral hydrogen particles and the possible truncating influence of self-absorption. Moreover, some profiles show additional intensity maxima, which cannot be derived from Eqs. (78) and (80). These maxima are clearly attributed to a local

increase in hydrogen density, while in the derivation of (80) we assumed constant hydrogen density throughout the entire space. The intensity maxima are often repeated in profiles corresponding to nearby values of l'. This shows that the regions of increased hydrogen density are fairly extended.

The hydrogen density distribution reproduced in Figure 163 was obtained by van de Hulst, Muller, and Oort from their line profiles in 54 directions in the longitude range $322° \leqslant l \leqslant 220°$. They also used galactic rotation data obtained from observations of stars, took into consideration the residual velocities of hydrogen particles and the dependence of intensity on particle concentration for a given v_r.

TABLE 7

l	l'	R_D, kpc	v_r at D, km/sec	v_c, km/sec	v_c', km/sec	ω, km/sec/pc	ω', km/sec/pc
45°.0	77°.5	8.01	+6	217	218	27.1	27.3
42 .5	75 .0	7.92	−3	212	219	26.8	27.8
41 .2	73 .7	7.87	+6	214	220	27.2	28.0
38 .4	70 .9	7.75	+8	213	221	27.5	28.5
35 .9	68 .4	7.62	+10	211	222	27.7	29.2
33 .4	65 .9	7.49	+16	214	223	28.6	29.8
30 .9	63 .4	7.33	+25	219	224	29.9	30.6
28 .4	60 .9	7.16	+29	218	225	30.4	31.4
25 .9	58 .4	6.98	+34	218	226	31.2	32.4
23 .4	55 .9	6.79	+41	220	226	32.4	33.3
20 .9	53 .4	6.58	+52	226	226	34.3	34.3
18 .4	50 .9	6.36	+55	223	226	35.1	35.5
15 .9	48 .4	6.13	+59	221	225	36.1	36.7
13 .4	45 .9	5.89	+62	217	224	36.8	38.0
10 .9	43 .4	5.63	+64	213	223	37.8	39.6
8 .4	40 .9	5.37	+67	209	221	38.9	41.2
5 .9	38 .4	5.09	+85	219	219	43.0	43.0
3 .4	35 .9	4.81	+89	216	217	44.9	45.1
0 .9	33 .4	4.51	+94	213	215	47.2	47.8
358 .4	30 .9	4.21	+97	208	212	49.4	50.4
355 .9	28 .4	3.90	+97	200	209	51.3	53.6
353 .4	25 .9	3.58	+104	199	205	55.6	57.3
350 .9	23 .4	3.26	+102	188	201	57.7	61.7
348 .4	20 .9	2.93	+118	195	196	66.6	66.9
345 .9	18 .4	2.59	+122	190	190	73.4	73.4
343 .4	15 .9	2.25	+74 +141	133 200	182	59.1 88.9	80.9
340 .9	13 .4	1.90	+92 +145	142 195	174	74.7 102.6	91.6
338 .4	10 .9	1.55	+102 +158	142 198	164	91.6 127.7	105.8
335 .9	8 .4	1.20	+112 +164	144 196	152	120.6 163.3	126.7
333 .4	5 .9	0.84	+121 +170	143 192	137	170.2 228.6	163.1
330 .8	3 .3	0.47	+128 +182	140 194	111	297.9 412.8	236.2
328 .3	0 .8	0.11	+56	59	49	536.4	445.5
325 .8	358 .3	0.24	−78	84	80	350.0	333.3
323 .6	356 .1	0.56	−66	81	119	144.6	212.5
321 .1	353 .6	0.91	−91 −136	115 160	140	126.4 175.8	153.8

Kwee, Muller, and Westerhout focused their attention on the kinematic problem. From the profiles of Figure 12 they determined the v_r corresponding to maximum intensity (presumably the point D). At this point maximum v_r is expected, but as we have noted above this is not quite so owing to the presence of additional factors. Kwee, Muller, and Westerhout took $l_0 = 327°.5$, $R_0 = 8200$ pc. Now, if v_r for the point D is known, $\omega(R)$ at that point (where $R = R_0 \sin l'$) is found from Eq. (78). The results are listed in Table 7.

The third column gives the distance of D from the center of the Galaxy, the fourth column lists the v_r at that point, the fifth the linear velocity of the

centroid, and the seventh gives the angular velocity. For $3°.3 \ll l' \ll 15°.9$ and $l' = 353°.6$, the last three parameters are determined with high uncertainty, and only the limits are indicated. Columns six and eight give the linear circular velocity and the angular velocity obtained after smoothing. These are currently the most reliable galactic rotation data; they are presented graphically in Figures 177 and 178, which plot the linear and the angular rotation velocity of hydrogen as a function of the distance from the center of the Galaxy. The smoother curve is drawn through the upper points, since in some cases the maximum intensity is received not from D but from other points where the hydrogen intensity is for some reason much higher than at D. This effect invariably lowers the values of ω and θ_c.

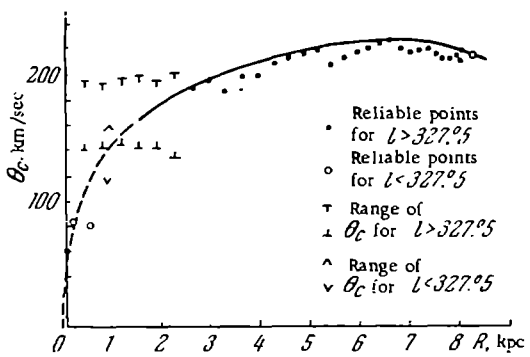


FIGURE 177

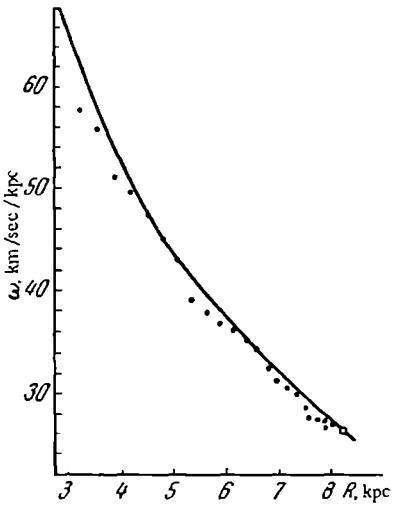


FIGURE 178

Note the high angular velocity of the central regions of the Galaxy. At a distance of 3 kpc from the center the angular velocity is four times that in the neighborhood of the Sun.

On passing from directions with galactic longitudes greater than the longitude of the center of the Galaxy to directions with longitudes less than that of the center, we expect the shift of the line profiles to change from positive to negative v_r. This is indeed observed on passing from $l = 328°.3$ to $l = 325°.8$. The currently accepted longitude of the galactic center 327°.0 (see §131) falls halfway between these values of l.

Of considerable interest are the wide and gentle wings of line profiles observed in the directions close to the galactic center (Kwee, Muller, and Westerhout). The profile in the direction $l = 321°.1$ has only a left-hand wing, and profiles in the directions $l = 330°.8$, 333°.4, 335°.9, 338°.4, 340°.9 and 343°.4 have only right-hand wings. The directions nearest the center $l = 323°.6$, 325°.8, 328°.3 have wings on both sides; for the first two directions the left wing is apparently the larger, and for the third direction the right wing is more prominent than the left. This phenomenon can be interpreted as indicating radial motion of hydrogen with velocities

of 50—100 km/sec from the central regions of the Galaxy. According to Oort and Rougoor, the total quantity of gas flowing out of a region of 3 kpc radius around the center of the Galaxy is about 1.5 solar masses annually.

§153. The spin of the stars. Let the star spin on its axis, which makes an angle i (other than zero) with the line of sight. Then the points on one half of the observed surface approach the observer, whereas the points of the other half recede from the observer. The resulting Doppler effect will broaden the lines in the spectra of these stars. Line broadening is also caused by other factors, e.g., motion of large gas clouds in stellar atmospheres (again a variety of the Doppler effect) and the influence of electric and magnetic fields on atomic radiation. However, if two stars of the same spectral type and the same luminosity show spectral lines of different widths, the difference can be naturally attributed to a difference in rotation velocities and the angle i (all other factors being taken equal). Under this assumption, the difference in the line widths for two such stars is clearly equal to the difference in the values of the function

$$y = x \sin i, \tag{81}$$

where x is the rotational velocity of points on the star's equator. This x is called the true rotation velocity of a star, and y is the apparent rotation velocity. Both are measured in km/sec.

If a fairly representative sample of stars of one spectral type is taken, we can say with fair certainty that those with the narrowest spectral lines have y equal virtually to zero (mainly because of small i and partly because of small x). Subtracting the line width of these stars from the line width of some other star, we obtain the apparent rotation velocity of the latter. The spin can also be determined from an examination of other factors causing line broadening, since in this way we can isolate the contribution of rotation velocity to line width.

Rotation velocities of stars were determined by Struve, Shain, Westgate, Slettebak, Huang Su-Shu, and others.

The data on hand show that the spin is sensitive to the spectral type. This is clear from Table 8, which is compiled from the data of Huang Su-Shu (1953). The table lists the spectral type, the number of stars investigated N, and the apparent rotation velocity $\bar{y}$.

TABLE 8

Sp	N	$\bar{y}$, km/sec	Sp	N	$\bar{y}$, km/sec
O	13	176	A4 – A5	102	120
B0 – B1	45	141	A7 – A9	31	112
B2 – B3	200	178	F0 – F1	91	89
B4 – B5	104	166	F2 – F3	52	57
B6 – B7	34	113	F4 – F5	99	44
B8 – B9	224	172	F6 – F7	36	18
A0 – A1	241	153	F8 – F9	35	9
A2 – A3	224	124			

The averaged variation of the rotation velocity with spectral types is illustrated by the spectrum — spin diagram shown in Figure 179.

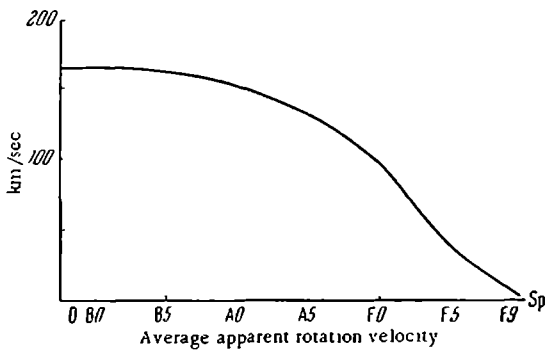


FIGURE 179

We see from the table that between O and B9 the average rotation velocity is virtually constant; the observed fluctuations are clearly of random origin, and are associated mainly with random fluctuations in the distribution of i. However, for spectral types later than A0, a steady and rapid decrease of rotation velocity is observed. For G0 stars the rotation velocity is so small that the corresponding broadening of spectral lines is undetectable. This is consistent with the known rotational velocity of the Sun (a G2 star), which is approximately 2 km/sec.

Kopylov and Boyarchuk compiled all the published data and plotted separate spectrum—spin diagrams for stars of luminosity classes I, II, III, IV, and V. The decrease of rotation velocities on passing from A0 to later spectral types was observed for all luminosity classes without exception.

The spin of a star may change in the course of its evolution for two reasons: change of radius and loss of mass. If the star undergoes isotropic expansion and its radius is increased a factor of k (the mass remaining constant), the moment of inertia relative to the spin axis is increased a factor of k^2, and since the total angular momentum is conserved the angular velocity will fall a factor of k^2, the rotation velocity decreasing a factor of k.

If the mass of the star decreases (individual atoms leaving the atmosphere through thermal dissipation, ejection of mass by powerful prominences, not unlike those on the Sun but larger in the early type stars), the star will gradually lose angular momentum (assuming that the mass is ejected at different directions and at different velocities). This is so because mass ejected with individual velocity close to the direction of the star's spin is more likely to escape into outer space than mass ejected in other directions. The angular momentum of these ejections is greater than the average momentum of stellar matter of the same mass, and the phenomenon therefore involves irreversible loss of angular momentum leading to a deceleration of the star's spin. Calculations indicate that if stellar evolution involves loss of mass, this effect is quite significant and may acutally account for the shape of the spectrum—spin diagram.

Note that the observations give the apparent rotation velocity y, whereas it is the true rotation velocity x that we are interested in. Chandrasekhar and Munch (1950) tried to recover the frequency function $f(x)$ from the observed distribution $\varphi(y)$. It is advisable to solve this problem separately

for stars of each spectral type (if N is sufficiently large). The analysis is based on the assumption that the various directions of the spin axis are equiprobable, and the frequency of i is thus equal to $\sin i$.

Since x and i are independent variables, the probability that they occur jointly in specified elementary intervals is given by the product

$$f(x)\sin i\,dx\,di. \tag{82}$$

Consider some fixed interval $[y, y+dy]$. The probability for the apparent rotation velocity to fall in this interval can be found by changing over in (82) from i to x and y using (81) (y is fixed!) and then integrating over all $x \geqslant y$:

$$\varphi(y)\,dy = dy\int_y^\infty \frac{y f(x)}{x\sqrt{x^2-y^2}}\,dx. \tag{83}$$

This integral equation can be reduced to Abel's equation and solved analytically for the sought function $f(x)$:

$$f(x) = \frac{2}{\pi x}\frac{d}{dx}\int_x^\infty \frac{\varphi(y)}{\sqrt{y^2-x^2}}\,dy.$$

In practice, however, this solution is not very useful. The main thing is not to find the function $f(x)$ but to devise a simple method for the calculation of the moments of this function. To this end, we multiply the two sides of Eq. (83) by y^n, integrate from 0 to ∞, and then change the order of integration in the right-hand side of the equation:

$$\int_0^\infty y^n\varphi(y)\,dy = \int_0^\infty \frac{1}{x}f(x)\,dx\int_0^x \frac{y^{n+1}}{\sqrt{x^2-y^2}}. \tag{84}$$

Changing over to a new variable of integration $t = \frac{y^2}{x^2}$, we get

$$\int_0^\infty y^n\varphi(y)\,dy = \frac{1}{2}\int_0^\infty x^n f(x)\,dx\int_0^1 t^{\frac{n}{2}}(1-t)^{-\frac{1}{2}}\,dt \tag{85}$$

The integral over t is the beta function $B\left(\frac{n}{2}+1, \frac{1}{2}\right)$. Using the standard ralation between beta and gamma functions, we finally get

$$\overline{y^n} = \frac{1}{2}\sqrt{\pi}\,\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n}{2}+\frac{3}{2}\right)}\,\overline{x^n}, \tag{86}$$

where $\overline{x^n}$ and $\overline{y^n}$ are the average values of x^n and y^n.

From (86), in particular, we have

$$\overline{x} = \frac{4}{\pi}\overline{y}, \tag{87}$$

$$\overline{x^2} = \frac{3}{2}\overline{y^2}, \tag{88}$$

$$\overline{x^3} = \frac{16}{3\pi}\overline{y^3}, \tag{89}$$

and also

$$\overline{(x-\bar{x})^2} = 1.5\,\bar{y}^2 - \frac{16}{\pi^2}\,\bar{y}^2, \tag{90}$$

$$\overline{(x-\bar{x})^3} = \frac{16}{3\pi}\,\bar{y}^3 - \frac{18}{\pi}\,\bar{y}^2\bar{y} + \frac{128}{\pi^3}\,\bar{y}^3. \tag{91}$$

Eq. (87) enables us to change over from the apparent rotation velocity to the average true rotation velocity, and Eq. (90) gives the dispersion of true rotation velocities. From Eq. (90) we see that if the dispersion of the true rotation velocities is zero, the dispersion of the apparent rotation velocities caused by different orientations of the spin axes is expressed by

$$\bar{y}^2 - \bar{\bar{y}}^2 = \left(\frac{32}{3\pi^2} - 1\right)\bar{y}^2. \tag{92}$$

and the root mean square deviation of the average apparent rotation velocity for N stars is

$$\frac{1}{\sqrt{N-1}}\sqrt{\bar{y}^2 - \bar{\bar{y}}^2} = \sqrt{\frac{32}{3\pi^2} - 1}\,\frac{\bar{y}}{\sqrt{N-1}} \approx \frac{0.28}{\sqrt{N-1}}\,\bar{y}. \tag{93}$$

The last expression can be applied to analyze the figures in the table. For spectral types B6—B7 we get $\frac{1}{\sqrt{N-1}}\sqrt{\bar{y}^2 - \bar{\bar{y}}^2} \approx 5.5$ km/sec. This is clearly too low to fully account for the minimum of the average apparent rotation velocity observed in this part of the table. On the other hand, such a narrow minimum can hardly be regarded as statistically significant. We therefore conclude that the dispersion of the true rotation velocities in this particular narrow band of spectral types is also substantial. Moreover, the deviations of $\bar{y}$ increase due to errors in the determination of the apparent rotation velocities.

REFERENCES TO CHAPTER XIX

1. DEICH, A.N. Secular parallaxes of faint stars, derived from the Pulkovo Catalogue of proper motions in Kapteyn's areas.— Izvestiya, Pulkovo Obs., No. 138. 1947.
2. SONN, W. and K. RUDNITSKI. Stellar Astronomy.— [Russian translation. 1959.]
3. OGORODNIKOV, K.F. Elements of the dynamics of rotating stellar systems.— Uspekhi Astron. Nauk, 4:3. 1948.
4. OGORODNIKOV, K.F. Dinamika zvezdnykh sistem (Dynamics of Stellar Systems).— Fizmatgiz. 1958.
5. PARENAGO, P.P. A study of spatial velocities of stars.— Astron. Zh., 27:150. 1950.
6. PARENAGO, P.P. Kurs zvezdnoi astronomii (A Course in Stellar Astronomy).— Gostekhizdat. 1954.
7. PARENAGO, P.P. On the division of the main sequence into two subsequences.— Voprosy Kosmogonii, 2:261. 1954.
8. SOBOLEV, V.V. Derivation of the spatial distribution of star velocities from proper motions.— Astron. Zh., 15:442. 1938.
9. AMBARZUMIAN, V.A. On the derivation of the frequency function of space velocities of the stars from the observed radial velocities.— MN, 96, 172, 1936.
10. CAMM, G.L. A study of galactic rotation based on the velocities of the planetary nebulae.— MN, 99:71. 1939.
11. CAMPBELL, W.W. Stellar motions. 1913.

12. EDDINGTON, A.S. Stellar Movements and the Structure of the Universe. 1914.
13. CHANDRASEKHAR, S. and G. MUNCH. On the integral equation governing the distribution of the true and the apparent rotational velocities of stars — Astrophys. J., 111:142. 1950.
14. KWEE, K.K., C.A. MULLER and G. WESTERHOUT. The rotation of the inner parts of the galactic system.— Bull. Astron. Inst. Netherl., 12, No.458. 1954.
15. OORT, J.H. Dynamics of the galactic system in the vicinity of the Sun.— Bull. Astron. Inst. Netherl., 4:269. 1928.
16. OORT, J.H. Mean parallaxes of faint stars derived from the Radcliffe Catalogue of proper motions.— Bull. Astron. Inst. Netherl., 8, No.290. 1936.
17. PAHLEN, E. Lehrbuch der Stellarstatistik. Leipzig.
18. SMART, W.M. Stellar Dynamics — Cambridge, 1938.
19. SCHWARZSCHIELD, K. Zur Stellarstatistik.— Astron. Nachr., 190:361. 1912.
20. HUANG SU-SHU. A statistical study of the rotation of the stars.— Astrophys. J., 118:285. 1953.
21. TRUMPLER, B.R. and H. WEAVER. Statistical Astronomy.— California, USA. 1953.

T. A. AGEKYAN

Chapter XX

STELLAR DYNAMICS

§154. Basic premises. Stellar dynamics is concerned with the study of the regular features of motion in star systems and the effect of this motion on the evolution of the systems. In this context a star system (or stellar system) is not only a galaxy but in fact any system having stars as its main constituent, e.g., globular star clusters, open clusters, or multiple stars. Until recently, stellar dynamics was mainly confined to a study of the stellar component of systems, and it is this branch of stellar dynamics which has reached maximum development. Lately, however, it has become increasingly clear that the dynamics of certain classes of star system is largely influenced by dust and gaseous matter as well.

Gravitation is the only force governing the interaction of stars within the framework of stellar dynamics. Radiation pressure is significant in the treatment of star interactions with diffuse matter. Finally, the motion of clouds of diffuse matter far from high-luminosity stars is largely determined by the magnetic fields in the interstellar space.

The gravitational field of a star system has a highly complex geometry. Since the gravitational force of a point mass is inversely proportional to the distance squared — this is not a particularly rapidly decreasing function — the overall gravitational force of the constituents throughout most of the stellar system is substantially greater than the gravitational force of that one object which makes the maximum contribution to the resultant at the particular point. Whenever this is true, the gradient of the gravitational field is obviously small.

On the other hand, in the immediate vicinity of stars, dense stellar clusters, and other compact objects, the gravitation of the nearby object is comparable with the overall gravitation of all the other objects, or is even greater. Here the gravitational force is a rapidly varying function of position. Following Ambartsumyan, we thus introduce the regular and the irregular gravitational field in stellar system.

The regular field is the field set up by an artificially smoothened uniform distribution of matter in the system. The forces corresponding to this field are called regular forces.

The irregular field is the field derived as the difference between the actual field of a system and its regular field. The corresponding forces are naturally called irregular forces.

The irregular field is a direct consequence of the discrete structure of stellar systems. The irregular forces on a body moving in a stellar system

are in fact random forces. Hence, the irregular force which a moving body experiences in an encounter with some object in the system is independent of the irregular force that the same body experienced in a previous encounter with some other object in the system.

The gravitational field of a star system is thus characterized, on the one hand, by its continuity and, on the other, by its discrete structure.

The larger the population of a star system, the greater is the contribution from the regular forces to its dynamics and the less significant are the irregular forces.

This can be stated in general terms as follows: in systems of progressively more uniform structure the role of the regular forces increases and the significance of the irregular forces diminishes.

This proposition can be demonstrated by a simple calculation. Let us find the ratio of the volume where the irregular forces dominate to the entire volume of the system. For simplicity consider a spherical system with uniform distribution of stars. Let N be the number of objects in the system, ϱ its radius, m the mass of an individual object. Under these conditions, the gravitational force per unit mass at a point inside the system is

$$\frac{GN\bar{m}}{\varrho^3} r, \tag{1}$$

where G is the gravitational constant, r is the distance from the center of the system, $\bar{m}$ is the mean stellar mass.

On the other hand, the force of attraction per unit mass due to the nearest object is

$$\frac{Gm}{r_1^2}, \tag{2}$$

where r_1 is the distance from the object.

Equating (1) and (2) we find the radius of the sphere inside which the irregular force due to the individual object is greater than the regular force. The corresponding spherical volume is

$$\frac{4}{3}\pi\left(\frac{m\varrho^3}{N\bar{m}r}\right)^{\frac{3}{2}}. \tag{3}$$

Averaging (3) over the masses and dividing by the volume per star $\left(\frac{4}{3}\frac{\pi\varrho^3}{N}\right)$, we obtain the sought ratio for a volume element:

$$\frac{1}{\sqrt{N}}\frac{\overline{m^{\frac{3}{2}}}}{(\bar{m})^{\frac{3}{2}}}\left(\frac{\varrho}{r}\right)^{\frac{3}{2}}. \tag{4}$$

Integration of (4) over the sphere gives, after dividing by the total volume, the sought ratio for the entire system:

$$\frac{2}{\sqrt{N}}\frac{\overline{m^{\frac{3}{2}}}}{(\bar{m})^{\frac{3}{2}}}. \tag{5}$$

The ratio (5) shows that in systems with N of the order of 10 (e.g., multiple stars), the volume of the irregular field is comparable with the volume of the regular field. In open clusters, with some 100 stars, the irregular field still occupies a substantial proportion of the entire volume of the star system. In galaxies (if these are considered as systems composed of individual stars) N is of the order of $10^{10}-10^{11}$ and the irregular field occupies a vanishingly small fraction of the total system volume. If, however, the galaxies are considered as systems composed of large objects—star clouds, globular clusters, groups of diffuse clouds, etc., N is found to be 5—7 orders of magnitude less and the proportion of the irregular field markedly increases. Expression (5) further shows that for a given N a large dispersion in the constituent masses increases the proportion of the irregular field in the system.

Galaxies and other stellar systems contain a large stellar population. They can thus be treated by the methods of statistical mechanics, which ignore the individual behavior of each particle in the system and concentrate on the statistical parameters which characterize the entire system as one whole. Stellar systems can be considered as composed of a very tenuous gas where the molecules move in the force field set up by the gas itself. Thus, globular clusters are an analog of a free gas sphere.

On the other hand, the trajectories (orbits) of the individual bodies in the stellar system, e.g., the orbits of the Sun or other stars in the Galaxy, are also of considerable interest. In solving these problems, the methods of stellar dynamics naturally approach the standard methods of celestial mechanics.

Any stellar system, like ordinary free gas with negative total energy, evolves toward an equilibrium figure. As long as the steady-state, equilibrium configuration has not been attained, the stellar system is said to be nonstationary.

We must make a distinction, however, between the equilibrium (stationary) state in relation to the regular forces alone and the equilibrium (stationary) state in relation to both the regular and the irregular forces. In the former case, the system parameters do not change in the regular field, where detailed balance has been established. In other words, if the irregular forces are ignored, the number of particles in any volume element of the phase space remains constant. If now the number of particles in each phase volume element is constant under the combined action of regular and irregular forces, the system is in equilibrium in relation to both regular and irregular forces. A nonstationary system approaches equilibrium under the combined action of regular and irregular forces, so that it becomes stationary in relation to both force fields. Under certain special conditions, however, the equilibrium in relation to regular forces may take a substantially shorter time to attain. Such a system, having reached an equilibrium in relation to regular forces only, will evolve further approaching equilibrium in relation to both regular and irregular forces.

The state of perfect equilibrium is never attained, however. After all, some of the bodies can be accelerated to supercritical velocities by the irregular field (e.g., following close encounters with other bodies in the system) and thus escape from the system. This dissipation alters the system parameters and the equilibrium is broken. In high-population systems the relative dissipation is slow so that the system has enough time to adjust to a new state of equilibrium, which is however promptly broken

by further dissipation. The combined effect of irregular and regular forces thus marches the system through a succession of equilibrium states. Ambartsumyan referred to this as a quasistationary state.

§155. The fundamental equation of stellar dynamics for the regular field. At this stage, we consider the regular field only. Let q_1, q_2, q_3 be the Hamiltonian (generalized) coordinates of a star and p_1, p_2, p_3 the corresponding momenta. Each star is described by a point $(q_1, q_2, q_3, p_1, p_2, p_3)$ in the six-dimensional phase space, and the entire stellar system is represented by the set of image points in the phase space. If $H(p, q)$ is the Hamiltonian, the equations of motion of a star take the form

$$\left.\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i}, \\ i &= 1, 2, 3. \end{aligned}\right\} \tag{6}$$

Let $\psi(q_1, q_2, q_3, p_1, p_2, p_3, t)$ be the density function in the phase space, and V the generalized velocity vector in the phase space, with the components $\dot{q}_1, \dot{q}_2, \dot{q}_3, \dot{p}_1, \dot{p}_2, \dot{p}_3$.

We can thus write the equation of continuity, which is an expression of the constancy of the total number of particles (points in the phase space):

$$\frac{\partial \psi}{\partial t} + \operatorname{div}(\psi V) = 0. \tag{7}$$

If the system is stationary, ψ is not an explicit function of time, so that $\frac{\partial \psi}{\partial t} = 0$.

Eq. (7) can be written in expanded form

$$\frac{\partial \psi}{\partial t} + \sum_{i=1}^{3} \left[\frac{\partial}{\partial q_i}(\psi \dot{q}_i) + \frac{\partial}{\partial p_i}(\psi \dot{p}_i) \right] = 0,$$

or, developing the derivatives,

$$\frac{\partial \psi}{\partial t} + \sum_{i=1}^{3} \left[\dot{q}_i \frac{\partial \psi}{\partial q_i} + \dot{p}_i \frac{\partial \psi}{\partial p_i} \right] + \psi \sum_{i=1}^{3} \left[\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right] = 0. \tag{8}$$

By (6)

$$\frac{\partial \dot{q}_i}{\partial q_i} = -\frac{\partial \dot{p}_i}{\partial p_i} = \frac{\partial^2 H}{\partial q_i \partial p_i},$$

so that Eq. (8) takes the simpler form

$$\frac{\partial \psi}{\partial t} + \sum_{i=1}^{3} \left[\dot{q}_i \frac{\partial \psi}{\partial q_i} + \dot{p}_i \frac{\partial \psi}{\partial p_i} \right] = 0. \tag{9}$$

It can be written in an alternative form

$$\frac{D\psi}{Dt} = 0, \tag{10}$$

where $\frac{D}{Dt}$ is the Stokes operator, which stands for the rate of change of a function at a point moving in the phase space together with the particle.

Eq. (10), known as the Boltzmann equation, was first applied to stellar systems by Jeans. This is an expression of Liouville's theorem, which states that the density of a volume element in the phase space remains constant as it moves along a phase trajectory.

In rectangular coordinates, the Hamiltonian per unit mass is written in the form

$$H = \frac{1}{2}(u^2 + v^2 + w^2) - \Phi(x, y, z, t),$$

where u, v, w are the stellar velocity components, and $\Phi(x, y, z, t)$ is the regular gravitational potential. The canonical equations of motion (6) in rectangular coordinates thus take the form

$$\left.\begin{aligned} \dot{x} &= \frac{\partial H}{\partial u}, & \dot{y} &= \frac{\partial H}{\partial v}, & \dot{z} &= \frac{\partial H}{\partial w}, \\ \dot{u} &= -\frac{\partial H}{\partial x}, & \dot{v} &= -\frac{\partial H}{\partial y}, & \dot{w} &= -\frac{\partial H}{\partial z}. \end{aligned}\right\} \tag{11}$$

where

$$u = \frac{\partial \Phi}{\partial x}, \quad \dot{v} = \frac{\partial \Phi}{\partial y}, \quad \dot{w} = \frac{\partial \Phi}{\partial z}. \tag{12}$$

Eq. (9) in rectangular coordinates is written in the form

$$\frac{\partial \psi}{\partial t} + u\frac{\partial \psi}{\partial x} + v\frac{\partial \psi}{\partial y} + w\frac{\partial \psi}{\partial z} + \frac{\partial \Phi}{\partial x}\frac{\partial \psi}{\partial u} + \frac{\partial \Phi}{\partial y}\frac{\partial \psi}{\partial v} + \frac{\partial \Phi}{\partial z}\frac{\partial \psi}{\partial w} = 0. \tag{13}$$

Note that Eq. (13) is applicable only when the irregular forces are ignored. If the irregular field is also considered, the zero in the right-hand side of Eq. (13) is replaced by the function

$$\left(\frac{d\psi}{dt}\right)_{\text{irr}}. \tag{14}$$

which allows for the effect of irregular forces. By Lagrange's method, the partial linear differential equation (13) is equivalent to a system of ordinary differential equations,

$$dt = \frac{dx}{u} = \frac{dy}{v} = \frac{dz}{w} = \frac{du}{\frac{\partial \Phi}{\partial x}} = \frac{dv}{\frac{\partial \Phi}{\partial y}} = \frac{dw}{\frac{\partial \Phi}{\partial z}}, \tag{15}$$

and the general solution of Eq. (13) is an arbitrary function of six independent first integrals of (15).

Eqs. (15), however, are in fact the equations of motion (11), so that the first integrals of (15) are the first integrals of motion. Any combination of these integrals is clearly also an integral of motion. That the phase density ψ is a first integral of motion also follows directly from Eq. (10). Thus,

$$\psi = \psi(I_1, I_2, I_3, I_4, I_5, I_6),$$

where

$$I_i = I_i(t, x, y, z, u, v, w) \qquad (i = 1, 2, \ldots, 6),$$

are the first integrals of motion.

There is a further expression relating the functions ψ and Φ. If ψ is interpreted as the mass density in the phase space (and not the density of points), then integration of ψ over the velocity components gives the mass density in the physical space:

$$D = \int\int\int \psi \, du \, dv \, dw. \tag{16}$$

From the potential theory, the potential is related to the density by Poisson's equation:

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} = -4\pi G D. \tag{17}$$

Eqs. (16) and (17) provide the second relation between the potential and the phase density.

A system is stationary in a regular field if and only if it satifies Eq. (13) without the first term in the left-hand side — without the partial derivative of phase density with respect to time. By (16) and (17) the potential in this case is also independent of time.

A system is stationary in both the regular and the irregular field if and only if it satisfies Eq. (13) without the first term in the left-hand side and moreover the function (14) identically vanishes.

If the system is stationary in the regular field, the first equation in (15) drops out and the system reduces to the differential equations of the phase path. In a stationary system the phase density is in general a function of the five (and not six) first integrals of motion:

$$\begin{aligned} &\psi = \psi(I_1, I_2, I_3, I_4, I_5), \\ &I_i = I_i(x, y, z, u, v, w), \quad i = 1, 2, \ldots, 5. \end{aligned} \tag{18}$$

these first integrals being independent of time.

Each of the integrals of motion should clearly satisfy Eq. (13), i.e.,

$$\frac{DI_i}{Dt} = 0. \tag{19}$$

If the system is stationary in the regular field, Eq. (19) is clearly solved by the function

$$I_1 = u^2 + v^2 + w^2 - 2\Phi.$$

Therefore the first integral

$$I_1 = V^2 - 2\Phi = \text{const}, \tag{20}$$

known as the energy integral (this is an expression of the fact that the total energy of a star is constant), is applicable to any stationary system.

Let us consider in more detail the expression of phase density in terms of the independent first integrals of motion, following the analysis of Kuzmin (1953).

Each of the first integrals (18) defines a certain fixed five-dimensional hypersurface in the phase space (in the case of a nonstationary system, these hypersurfaces in general move). The intersection of these five hypersurfaces defines a line which specifies the path of a point in the phase space. Depending on the actual form of the potential function, the path of the point

may be closed or open; it is readily seen that if the trajectory is closed in the phase space it is closed in the physical space too, and vice versa. In general the trajectory is not closed, and its turns fill a two- or three-dimensional domain in the physical space, which corresponds to a hypersurface of two or more dimensions in the phase space. It is clear, however, that five five-dimensional hypersurfaces defined by independent equations (18) give a two-dimensional hypersurface only if the folds of one of these five-dimensional hypersurfaces fill the entire six-dimensional space. The corresponding integral (18) is obviously infinite-valued. This infinite-valued integral cannot be used in the representation of the phase density function which is a priori single-valued.

Similarly, if the open trajectory of a point fills a three-dimensional region in the phase space, two of the integrals in (18) are infinite valued and cannot be used in the representation of the phase density function, etc.

Therefore, in the general case of a system which is stationary in the regular field, Eqs. (18) are written in the form

$$\begin{aligned} &\psi = \psi(J_1, J_2, \ldots, J_n), \\ &J_j = J_j(x, y, z, u, v, w), \quad j = 1, 2, \ldots, n \leqslant 5. \end{aligned} \tag{18*}$$

where J_j are the single-valued first integrals of motion. The remaining first integrals of motion

$$J_k = J_k(x, y, z, u, v, w), \quad k = n+1, n+2, \ldots, 5$$

are infinite-valued and do not participate in the representation of the phase density function.

When Eq. (13) is solved in practice, we follow one of the two alternative methods. One possibility is to specify the potential Φ, to determine the first integrals of motion from (15) and thus find the properties of the phase density ψ. This was the method adopted by Jeans. The other possiblity is to specify the function ψ, insert it in (13) and thus derive relations between the parameters of the particular distribution used. This was the approach favored by Oort and Chandrasekhar.

§156. **The solution of the fundamental equation of stellar dynamics using a known potential.** We will first consider the application of Jeans's method. Consider a stationary system in the regular field with a known potential Φ. We start with the most general restrictions on the potential, which will be subsequently relaxed.

a) Let the potential be such that all the stars move in closed orbits. In mechanics there are only two particular forms of potential of this kind,

$$\Phi = c_1 r^{-1}, \tag{21}$$

$$\Phi = c_2 r^2 + c_3, \tag{22}$$

which ensure closed orbits for any x, y, z and u, v, w. (If in case (21) the velocity of the body is supercritical, the trajectory is open, but in a different sense. It does not fill a two- or three-dimensional region in space.) Here c_1, c_2, c_3 are constants, and r is the distance from the point O. The potentials (21) and (22) are particular cases of a general spherically symmetric potential.

We know from celestial mechanics that in case (21) the object moves in an ellipse with its focus at the point O, and in case (22) in an ellipse with

its center at the point O. In either case the orbits are defined by five parameters, the orbital elements. Since the orbital elements remain constant in all stages of motion, they are in fact independent first integrals of motion. Hence, the phase density is a function of five independent integrals of motion and has apparently the most general form. Examination of Poisson's equation (17) shows, however, that potential (21) is permissible only if the gravitational field is set up by a single point mass, which is inapplicable to stellar systems. Similarly (22) corresponds to the particular case of a system of uniform density, which though permissible in principle is highly restrictive. In this case all the orbits are necessarily circular, which greatly limits the generality of the phase density. The generality of the phase density is thus self-contradictory.

b) Let us consider the general case of a spherically symmetric potential

$$\Phi = \Phi(r) = \Phi\left(\sqrt{x^2+y^2+z^2}\right).$$

The trajectory of the body in this case lies in a plane through the center and the initial velocity vector. In general, however, the trajectory is no longer closed and its turns fill a ring, i.e., a two-dimensional region, in the orbital plane. In the phase space the path thus fills at least a two-dimensional region. Hence it follows that at least one of the first integrals of motion is infinite-valued and the phase density depends at most on four independent first integrals of motion. One of these is the energy integral (20):

$$J_1 = V^2 - 2\Phi.$$

Three other integrals are found by writing Eqs. (15) in the form

$$\frac{dx}{u} = \frac{dy}{v} = \frac{dz}{w} = \frac{du}{\frac{x}{r}\frac{\partial\Phi}{\partial r}} = \frac{dv}{\frac{y}{r}\frac{\partial\Phi}{\partial r}} = \frac{dw}{\frac{z}{r}\frac{\partial\Phi}{\partial r}}.$$

Hence, e.g.,

$$v\,dx = u\,dy,$$
$$x\,dv = y\,dy.$$

Adding up these equations and integrating, we find J_2. Similarly for J_3 and J_4:

$$\left.\begin{aligned} J_2 &= xv - yu = \text{const},\\ J_3 &= yw - zv = \text{const},\\ J_4 &= zu - xw = \text{const}. \end{aligned}\right\} \tag{23}$$

Eqs. (23) are called the areal integrals: they express the constancy of the projections of double the areal velocity of the body on the three coordinate planes.

We have thus obtained all the four single-valued integrals of motion, and we may write

$$\psi = \psi(J_1, J_2, J_3, J_4).$$

The three areal integrals (23) are clearly a consequence of the constancy of the areal velocity in the orbital plane, expressed by the single inequality

$$r^2(\vartheta^2 + P^2) = r^2(V^2 - \Pi^2) = \text{const}, \tag{24}$$

where Π, ϑ, and P are the orthogonal linear velocity components, corresponding to the spherical coordinates r, θ, and φ. The phase density is therefore a function of two integrals of motion only:

$$\psi = \psi[V^2 - 2\Phi,\ r^2(\vartheta^2 + P^2)]. \tag{25}$$

Eq. (24) can be obtained from Eqs. (16) and (17). It follows that as long as the potential Φ is a function of r alone, the dependence of the phase density on coordinates can be expressed in terms of r alone. Hence, the integrals J_2, J_3, and J_4 should enter ψ only as a certain combination, such that the dependence on the coordinates is expressed in terms of r alone. J_2, J_3, and J_4 are the projections of double the areal velocity in the orbital plane, (24), on the coordinate planes; the sought combination is therefore of the form

$$J_2^2 + J_3^2 + J_4^2 = r^2(\vartheta^2 + P^2),$$

which confirms the previous conclusion. Thus, because of the natural restriction imposed by Poisson's equation, the phase density in a spherically symmetric potential is a combination of only two integrals of motion of the form (25). Since the velocity components ϑ and P enter the phase density only as a combination $\vartheta^2 + P^2$, the solid of velocities at any point in the stellar system is a solid revolution with its axis coinciding with the radius-vector r.

c) Consider now the potential of an axisymmetric stationary system. It is of particular interest as most of the galaxies, our Galaxy included, are axisymmetric to first approximation. The trajectory of a star in general case is no longer planar: it fills a three-dimensional torus-shaped volume (circular, though with a non-circular cross section). Hence the phase velocity is a function of at most three independent first integrals of motion:

$$\psi = \psi(J_1, J_2, J_3).$$

One of the first integrals J_1 is again the energy integral (20). Another integral is found by directing the z axis along the symmetry axis of the potential, so that

$$\Phi = \Phi(\sqrt{x^2 + y^2},\ z).$$

It is easily verified that in this case only the first integral in (23) — the areal integral with respect to the z axis — remains valid:

$$J_2 = xv - yu = \text{const}.$$

We define cylindrical coordinates R, θ, Z and the corresponding orthogonal linear velocity components Π, ϑ, Z. The energy integral and the areal velocity integral are then written in the form

$$\left.\begin{aligned} J_1 &= \Pi^2 + \vartheta^2 + Z^2 - 2\Phi = \text{const}, \\ J_2 &= R\vartheta = \text{const}. \end{aligned}\right\} \tag{26}$$

The third integral cannot be derived unless additional restrictions are imposed on the potential. However, for any axisymmetric potential, we can select a so-called quasi-integral of motion, i.e., a particular combination

of coordinates and velocity components which is a very slow function of time. This quasi-integral should be introduced in the expression for the phase density. Indeed, if the phase density were a function of only two first integrals (26), the solid of velocities at any point of the stellar system would be a body of revolution about an axis along the velocity component ϑ; this is at variance with the previous result for a spherical potential, which is after all a particular case of an axisymmetric potential.

As we know, the solid of velocities in the Galaxy is definitely not a solid of revolution with an axis along ϑ. Moreover, observations show that the solid of velocities in the Galaxy is close to an ellipsoid whose axes in the directions Π and Z, far from being equal, show the maximum difference, being the major (the former) and the minor axis, respectively (note that the principal axes of the velocity ellipsoid obtained from observations do not quite coincide with the directions Π, ϑ, and Z). Therefore the third integral (or more precisely, quasi-integral) entering ψ should make a very substantial contribution, essentially destroying the symmetry of the distribution in Π and Z. Various forms of the third integral (quasi-integral) were proposed by Lindblad, Clark, and Kuzmin.

d) Finally let us consider a stationary system with an asymmetric potential. In general, there is only one integral of motion, the energy integral. If the phase density is a function of this one integral only,

$$\psi = \psi(V^2 - 2\Phi),$$

we see by (16) that the spatial density D is a function of the potential only,

$$D = D(\Phi)$$

and Poisson's equation (17), which is symmetric in x, y, z, inevitably gives a spherically symmetric potential Φ. We have shown above, however, that in a spherically symmetric potential the phase density is a function of two first integrals of motion. Hence, in a stationary system the phase density cannot be a function of one integral of motion only.

Reasoning by analogy with equilibrium configurations of a rotating self-gravitating liquid, we can assert that in general a stellar system which is stationary in a regular field should have a symmetry axis and a plane of symmetry.

§157. The hydrodynamic equations for stellar systems. Let us write Eq. (13) in cylindrical coordinates. This can be done applying the Hamiltonian expressed in cylindrical coordinates to Eq. (9) and then changing over from momenta to velocity components. A more educational approach, however, prescribes an independent derivation of Eq. (9) by an alternative technique, directly in cylindrical coordinates.

Let

$$\psi(R, \theta, z, \Pi, \vartheta, Z, t)\, dR\, d\theta\, dz\, d\Pi\, d\vartheta\, dZ\, dt \qquad (27)$$

be the number of stars with the coordinates R, θ, z and velocity components Π, ϑ, Z at the time t (to be more precise, we should say the number of stars with coordinates between R and $R + dR$, θ and $\theta + d\theta$, etc., at the time between t and $t + dt$, but we will nevertheless follow the shorter notation). After a time dt the coordinates and the velocity components of these stars change,

but since these are the same stars with new coordinates and velocity components, the function

$$\psi\left(R+\Pi\,dt,\ \theta+\frac{\vartheta}{R}\,dt,\ z+Z\,dt,\ \Pi+\dot{\Pi}\,dt,\ \vartheta+\dot{\vartheta}\,dt,\ Z+\dot{Z}\,dt\right)\times \\ \times\, dR\,d\vartheta\,dz\,d\Pi\,d\vartheta\,dZ\,dt \qquad (28)$$

is clearly equal to (27).

Equating the two expressions, expanding (28) in a Taylor series and retaining only terms of first order, we obtain on cancelling

$$\frac{\partial\psi}{\partial t}+\Pi\frac{\partial\psi}{\partial R}+\frac{\vartheta}{R}\frac{\partial\psi}{\partial\theta}+Z\frac{\partial\psi}{\partial z}+\dot{\Pi}\frac{\partial\psi}{\partial\Pi}+\dot{\vartheta}\frac{\partial\psi}{\partial\vartheta}+\dot{Z}\frac{\partial\psi}{\partial Z}=0. \qquad (29)$$

Lagrange's equations in cylindrical coordinates are written in the form

$$\left.\begin{aligned}\dot{\Pi}&=\frac{\vartheta^2}{R}+\frac{\partial\Phi}{\partial R},\\ \dot{\vartheta}&=\frac{1}{R}\frac{\partial\Phi}{\partial\theta}-\frac{\vartheta\Pi}{R},\\ \dot{Z}&=\frac{\partial\Phi}{\partial z},\end{aligned}\right\} \qquad (30)$$

where $\Phi(R, \theta, z, t)$ is the potential per unit mass. Eq. (29) is therefore written in the form

$$\frac{\partial\psi}{\partial t}+\Pi\frac{\partial\psi}{\partial R}+\frac{\vartheta}{R}\frac{\partial\psi}{\partial\theta}+Z\frac{\partial\psi}{\partial z}+\left(\frac{\vartheta^2}{R}+\frac{\partial\Phi}{\partial R}\right)\frac{\partial\psi}{\partial\Pi}+\left(\frac{1}{R}\frac{\partial\Phi}{\partial\theta}-\frac{\vartheta\Pi}{R}\right)\frac{\partial\psi}{\partial\vartheta}+\frac{\partial\Phi}{\partial z}\frac{\partial\psi}{\partial Z}=0. \qquad (31)$$

This is the fundamental equation of stellar dynamics in cylindrical coordinates. Successively multiplying Eq. (31) by

$$d\Pi\,d\vartheta\,dZ,\ \Pi\,d\Pi\,d\vartheta\,dZ,\ \vartheta\,d\Pi\,d\vartheta\,dZ,\ Z\,d\Pi\,d\vartheta\,dZ,$$

we integrate over the velocity components. Consider the following typical examples, which show how the terms of the resulting equations are transformed:

$$\int\limits_{-\infty}^{+\infty}\iiint\frac{\partial\psi}{\partial t}\,d\Pi\,d\vartheta\,dZ=\frac{\partial}{\partial t}\int\limits_{-\infty}^{+\infty}\iiint\psi\,d\Pi\,d\vartheta\,dZ=\frac{\partial}{\partial t}D,$$

$$\int\limits_{-\infty}^{+\infty}\iiint\frac{\partial\psi}{\partial t}\,\Pi\,d\Pi\,d\vartheta\,dZ=\frac{\partial}{\partial t}\int\limits_{-\infty}^{+\infty}\iiint\Pi\psi\,d\Pi\,d\vartheta\,dZ=\frac{\partial}{\partial t}D\overline{\Pi},$$

$$\int\limits_{-\infty}^{+\infty}\iiint\vartheta\Pi\frac{d\psi}{d\vartheta}\,d\Pi\,d\vartheta\,dZ=\frac{\partial}{\partial\vartheta}\int\limits_{-\infty}^{+\infty}\iiint\psi\vartheta\Pi\,d\Pi\,d\vartheta\,dZ=\frac{\partial}{\partial\vartheta}D\overline{\vartheta\Pi},$$

$$\int\limits_{-\infty}^{+\infty}\iiint\Pi\vartheta^2\frac{\partial\psi}{\partial\Pi}\,d\Pi\,d\vartheta\,dZ=\int\limits_{-\infty}^{+\infty}\iint[\psi\Pi]_{\Pi=-\infty}^{\Pi=+\infty}\vartheta^2\,d\vartheta\,dZ- \\ -\int\limits_{-\infty}^{+\infty}\iiint\psi\vartheta^2\,d\Pi\,d\vartheta\,dZ=-D\overline{\vartheta^2},$$

$$\int\limits_{-\infty}^{+\infty}\iiint\Pi\frac{\partial\psi}{\partial\vartheta}\,d\Pi\,d\vartheta\,dZ=\int\limits_{-\infty}^{+\infty}\iint[\psi]_{\vartheta=-\infty}^{\vartheta=+\infty}\Pi\,d\Pi\,dZ=0.$$

Here D is the spatial velocity; a bar on top indicates average value. In the fourth example, the first term in the by-parts integration vanishes when the velocity component goes to infinity, since the distribution function goes to zero faster than any negative power of the velocity component.

Applying similar transformations to all the terms of the four equations, we get

$$\left.\begin{aligned}
&\frac{\partial}{\partial t}D+\frac{\partial}{\partial R}(D\overline{\Pi})+\frac{1}{R}\frac{\partial}{\partial\theta}(D\overline{\vartheta})+\frac{\partial}{\partial z}(D\overline{Z})+\frac{1}{R}D\overline{\Pi}=0,\\
&\frac{\partial}{\partial t}(D\overline{\Pi})+\frac{\partial}{\partial R}(D\overline{\Pi^2})+\frac{1}{R}\frac{\partial}{\partial\theta}(D\overline{\Pi\vartheta})+\frac{\partial}{\partial z}(D\overline{\Pi Z})+\frac{1}{R}D(\overline{\Pi^2}-\overline{\vartheta^2})=D\frac{\partial\Phi}{\partial R},\\
&\frac{\partial}{\partial t}(D\overline{\vartheta})+\frac{\partial}{\partial R}(D\overline{\Pi\vartheta})+\frac{1}{R}\frac{\partial}{\partial\theta}(D\overline{\vartheta^2})+\frac{\partial}{\partial z}(D\overline{\vartheta Z})+\frac{2}{R}D\overline{\Pi\vartheta}=\frac{1}{R}D\frac{\partial\Phi}{\partial\theta},\\
&\frac{\partial}{\partial t}(D\overline{Z})+\frac{\partial}{\partial R}(D\overline{\Pi Z})+\frac{1}{R}\frac{\partial}{\partial\theta}(D\overline{\vartheta Z})+\frac{\partial}{\partial z}(D\overline{Z^2})+\frac{1}{R}D\overline{\Pi Z}=D\frac{\partial\Phi}{\partial z}.
\end{aligned}\right\}\qquad(32)$$

The differential equations (32) relating the average velocity components and their derivatives to the system potential are called the hydrodynamic equations of the stellar system. $\overline{\Pi}$, $\overline{\vartheta}$, and $\overline{Z}$ are the centroid velocity components. If the system is stationary and has an axis of circular symmetry, the partial derivatives with respect to t and θ vanish. If we further assume that ψ is an even function of Π and Z, which holds true in particular for an ellipsoidal velocity distribution, we get

$$\overline{\Pi}=\overline{Z}=\overline{\Pi\vartheta}=\overline{\vartheta Z}=\overline{Z\Pi}=0.$$

The first and the third equation in (32) then identically vanish, and the second and the fourth equations take the form

$$\frac{\partial}{\partial R}(D\overline{\Pi^2})+\frac{1}{R}D(\overline{\Pi^2}-\overline{\vartheta^2})=D\frac{\partial\Phi}{\partial R},\qquad(33)$$

$$\frac{\partial}{\partial z}(D\overline{Z^2})=D\frac{\partial\Phi}{\partial z}.\qquad(34)$$

Let the linear rotation velocity of the centroid be $\vartheta_0=\overline{\vartheta}$, and the residual velocity of the star in the sense of rotation $\vartheta'=\vartheta-\vartheta_0$. Then

$$\overline{\vartheta^2}=\overline{(\vartheta'+\vartheta_0)^2}=\overline{\vartheta'^2}+\vartheta_0^2.$$

Seeing that

$$\frac{\partial\Phi}{\partial R}=-\frac{1}{R}\vartheta_c^2,$$

where ϑ_c is the linear velocity of a point in circular orbit, we write (33) in the form

$$\vartheta_c^2-\vartheta_n^2=\overline{\vartheta'^2}-\overline{\Pi^2}-\frac{R}{D}\frac{\partial}{\partial R}(D\overline{\Pi^2}).\qquad(35)$$

Eq. (35) defines the difference between the circular velocity and the centroid velocity at a point. It shows, in particular, that as the residual velocities go to zero this difference also goes to zero in proportion to the square of the residual velocities. The centroids of a group of stars with a small dispersion of velocities move virtually with circular velocity. This is clearly valid only if the distribution function is indeed even with respect to Π and Z, as assumed above.

§158. The solution of the fundamental equation of stellar dynamics using a known velocity distribution function. Let us now consider the solution of Eqs. (13) for Φ assuming a known ψ. First we consider two particular cases.

a) Let the velocities of all stars be directed along the spherical coordinate r, so that at any point of the physical space the star may have only one of the two velocities $\Pi(r)$ or $-\Pi(r)$. In this case, if the stellar system is assumed stationary, the potential should be spherically symmetric with respect to the origin O. At any point the number of stars moving to the center (i.e., those with velocities $-\Pi(r)$) should be equal to the number of stars moving away from the center with velocities $\Pi(r)$. The number of stars crossing in unit time any spherical surface centered at O is therefore constant, so that we have

$$\Pi D r^2 = c, \tag{36}$$

which is clearly seen to be an integral of Eq. (13) for the particular case being considered. On the other hand, Poisson's equation (17) in spherical coordinates takes the form (note that because of the spherical symmetry of the potential its partial derivatives over spherical angles vanish)

$$\frac{d}{dr}\left(r^2 \frac{d\Phi}{dr}\right) = -4\pi G D r^2. \tag{37}$$

Further seeing that

$$\frac{d\Phi}{dr} = \frac{d\Pi}{dt} = \frac{d\Pi}{dr}\frac{dr}{dt} = \Pi \frac{d\Pi}{dr} = \frac{1}{2}\frac{d(\Pi^2)}{dr}$$

and transforming the right-hand side of (37) with the aid of (36), we obtain a second-order differential equation for Π^2:

$$\frac{d}{dr}\left[r^2 \frac{d(\Pi^2)}{dr}\right] = -\frac{8\pi G C}{\sqrt{\Pi^2}}. \tag{38}$$

This is an Emden-type equation. It cannot be solved analytically, but a numerical solution is obtained without difficulty. Judging by its structure we see that on specifying the initial condition $\Pi(0) = C_1 < C_2$, where C_2 depends on the value of C, we automatically define a constant b such that $\Pi \to 0$ for $r \to b - 0$.

From this result and (36) we see that in the system being considered the density at the center and on the surface of a sphere of radius $r = b$ is infifite. Inside the sphere the density is finite, and outside the sphere it is zero. The number of stars in the system is limited so that

$$\int_0^b 4\pi D r^2 dr = 4\pi \int_0^b \frac{dr}{\Pi(r)}$$

converges since $\Pi(r)$ goes very slowly to zero for $r \to b$. The existence of a stationary spherical stellar system with stars having pure radial velocities is thus permissible.

b) Now let the radial velocity component of all stars be zero, i.e., $\Pi = 0$. In a stationary system this is permissible only if the potential is spherically symmetric and the velocity $\sqrt{\Theta^2 + P^2}$ of each star satisfies the condition

$$\frac{\Theta^2 + P^2}{r} = \frac{GM}{r^2}.$$

where M is the mass of the system enclosed inside a sphere of radius r. Expressing M in terms of the density D, we find

$$\frac{\vartheta^2 + \Pi^2}{r} = \frac{4\pi G}{r^2} \int_0^r D(x)\, x^2\, dx. \tag{39}$$

It is essential that the velocity distribution at any point of the system be circularly symmetric in a plane tangent to the sphere at the particular point.

If the above conditions are satisfied, a stationary system with any given density function may exist.

It is quite obvious that the two extreme cases considered above correspond to a system which is stationary only with respect to regular forces. Encounters between stars should cause appearance and growth of velocity components perpendicular to the radius-vector in system (a) and appearance and growth of radial velocity components in system (b). This process eventually lead to establishment of a certain most probable velocity distribution in which the system remains stationary in relation to both the regular and the irregular forces. As we have already noted in §154, dissipation of stars from the system precludes establishment of perfect equilibrium: the system will attain a quasistationary state.

The above argument suggests the following three different approaches to the solution of Eq. (13) assuming a known function ψ:

1) The velocity distribution function is chosen proceeding from considerations associated with the origin of the stellar system, which is assumed to be young so that the irregular forces have not substantially modified the initial velocity distribution.

2) The velocity distribution function obtained from observations is used in the equations.

3) The system is assumed to have attained stationary state in relation to both types of forces and Eq. (13) is solved with function (14) inserted in the right-hand side.

Observations show (see §147) that the dispersions of residual velocities of stars in the directions Π, ϑ, and Z are not equal. This led Schwarzschild to the idea (by analogy with the Maxwellian velocity distribution, which is spherical) that the residual stellar velocities follow a power ellipsoidal distribution. Chandrasekhar, and later other authorities as well, considered a more general ellipsoidal distribution of residual velocities, which is used in the following treatment.

The centroids revolve in circular orbits around the symmetry axis of the Galaxy. In the cylindrical system of coordinates with the z axis directed along the rotation axis of the Galaxy and the principal plane coinciding with the galactic plane, the distribution function therefore has the form

$$\psi(R, \theta, z, \Pi, \vartheta, Z) = \psi(Q + g), \tag{40}$$

where

$$Q = h^2\Pi^2 + k^2(\vartheta - \vartheta_0)^2 + l^2Z^2 + mn\Pi(\vartheta - \vartheta_0) + n\Pi Z + p(\vartheta - \vartheta_0)z, \tag{41}$$

and the parameters g, h, k, l, m, n, p and ϑ_0 depend only on R, z and time; if the system is stationary in relation to regular forces, they only depend on R and z. ϑ_0 is the linear centroid velocity.

We will only consider a stationary system, inserting (40) in (31). Since (31) is linearly homogeneous with respect to partial derivatives of ψ, insertion of (40) followed by simple cancellations gives the same equations for the function $Q + g$. This conclusion also follows alternatively from the fact that since ψ should be a combination of the integrals of motion and depends only on $Q + g$, $Q + g$ itself should be a combination of the integrals of motion and thus satisfy Eq. (31). Using (41), we could insert $Q + g$ in the left-hand side of (31), carry out the differentiations, equate the coefficients of identical combinations of the powers of Π, ϑ, and Z and examine the resulting system of 21 equations, of which 12 are independent. This was the method used by Oort in 1927 in his solution of the problem. A simpler method is to represent $Q + g$ as a combination of the integrals of motions that we have obtained for an axisymmetric system:

$$\left.\begin{aligned} I_1 &= \Pi^2 + \vartheta^2 + Z^2 - 2\Phi, \\ I_2 &= R\vartheta. \end{aligned}\right\} \tag{42}$$

Since $Q + g$ is a polynomial of second degree in Π, ϑ and Z, I_1 is also a polynomial of second degree, and I_2 is a polynomial of first degree, we naturally seek a combination of the form

$$Q + g = c_1 I_1 + c_2 I_2^2 + 2c_3 I_2. \tag{43}$$

where c_1, c_2, and c_3 are constant coefficients. Inserting (41) and (42) in (43) and equating the coefficients of identical combinations of the powers of Π, ϑ, and Z in the left- and right-hand sides of the equation, we find

$$\left.\begin{aligned} m &= 0, \\ n &= 0, \\ p &= 0, \end{aligned}\right\} \tag{44}$$

$$h^2 = l^2 = c_1, \tag{45}$$

$$k^2 = c_1 + c_2 R, \tag{46}$$

$$\vartheta_0 = \frac{c_3 R}{c_1 + c_2 R^2}, \tag{47}$$

$$g = -2c_1\Phi - c_3^2 R^2. \tag{48}$$

Eqs. (44) show that (41) is of the form

$$Q = h^2\Pi^2 + k^2(\vartheta - \vartheta_0)^2 + l^2 Z^2. \tag{49}$$

i.e., the axes of the velocity ellipsoid point along the velocity components in the cylindrical frame.

Writing Eq. (47) in the form

$$\omega = \frac{c_3}{c_1 + c_2 R^2}, \tag{50}$$

we see that for small R, specifically for $R^2 \ll \left|\frac{c_1}{c_2}\right|$, we are virtually dealing with rigid-body rotation, whereas for large R ($R^2 \gg \left|\frac{c_1}{c_2}\right|$) the angular velocity decreases in inverse proportion to the square of the distance.

Since Oort's coefficients (see Eq. (66), Ch. XIX) have the form

$$A = -\frac{1}{2} R \frac{d\omega}{dR}, \qquad B = A - \omega, \tag{51}$$

using (50) we find

$$\frac{-B}{A-B} = \frac{C_1}{C_1 + C_2 R^2},$$

and from (45) and (46)

$$\frac{h^2}{k^2} = \frac{-B}{A-B}, \tag{52}$$

a relation first derived by Lindblad. The coefficients h, l, and k are inversely proportional to the semiaxes of the velocity ellipsoid and therefore, writing σ_R, σ_θ, and σ_z for the semiaxes, we have

$$\sigma_R = \sigma_z, \tag{53}$$

$$\frac{\sigma_\theta}{\sigma_R} = \sqrt{\frac{-B}{A-B}}. \tag{54}$$

Thus, in an axisymmetric stationary system with ellipsoidal velocity distribution, the principal axes of the ellipsoid coincide with the velocity components in cylindrical coordinates, the semiaxes in the directions R and Z are equal in magnitude and the ratio of the semiaxis in the direction θ to the other two semiaxes is defined by relation (54). Eq. (48) written in the form

$$g = -\frac{1}{\sigma_R^2}\Phi - k^2\sigma_\theta^2 \tag{55}$$

is an expression of the relation among stellar density, potential, and the radial semiaxis of the velocity ellipsoid.

If for the neighborhood of the Sun we adopt Parenago's values of the coefficients $A = +0.0195$ km/sec/pc, $B = 0.013$ km/sec/pc, we find $\sigma_\theta/\sigma_R = 0.63$, in good agreement with the results of direct reduction of stellar velocity observations.

The observations, at variance with the above theory, give

$$\frac{\sigma_z}{\sigma_R} \approx 0.50,$$

i.e., the semiaxis σ_z is not one of two equal major semiaxes, but is the smallest of the three. We have previously commented on this fact in §154, where we have shown that if ψ is regarded as a function of only two integrals of motion I_1 and I_2, the velocity distribution function is symmetric in Π and Z. We have just shown that for an ellipsoidal velocity distribution ψ is a combination of I_1 and I_2, and the theory therefore naturally leads to equality of σ_R and σ_z. Since this result is at variance with observations, the distribution function of the residual velocities in the Galaxy of necessity depends on three independent integrals of motion, I_1, I_2, and I_3 (the last being a quasiintegral) and the actual distribution of residual velocities in the Galaxy cannot be ellipsoidal.

It will be shown in §159 that the distribution of the components of the residual velocity projections on the plane $z = 0$ in the region of circular orbits is determined by the differential rotation of the stellar system. In principle, the distribution is not elliptic, although in practice it is close to an elliptic distribution.

Note that the constancy of σ_R in a stationary stellar system with an ellipsoidal velocity distribution, which follows from (45), can be regarded as an indication of the fact that in the real distribution the dispersion of Π, i.e., $\overline{\Pi^2}$, is not particularly sensitive to R. In this case, Eq. (35) can be written in the form

$$\vartheta_c^2 - \vartheta_0^2 = \overline{\vartheta'^2} - \overline{\Pi^2}\left(1 + R\frac{\partial \ln D}{\partial R}\right). \tag{56}$$

Let us derive another relation between the kinematic coefficients and the density. We write Poisson's equation (17) in cylindrical coordinates seeing that $\frac{\partial \Phi}{\partial \theta} = 0$.

$$\frac{\partial^2 \Phi}{\partial R^2} + \frac{1}{R}\frac{\partial \Phi}{\partial R} + \frac{\partial^2 \Phi}{\partial z^2} = -4\pi G D. \tag{57}$$

Since

$$\frac{\partial \Phi}{\partial R} = -\omega^2 R,$$

we get

$$\frac{\partial^2 \Phi}{\partial R^2} + \frac{1}{R}\frac{\partial \Phi}{\partial R} = -2\omega\left(R\frac{\partial \omega}{\partial R} - \omega\right) = -2(A-B)(-2A+A-B) = 2(A^2 - B^2),$$

so that

$$4\pi G D = -\frac{\partial^2 \Phi}{\partial z^2} - 2(A^2 - B^2). \tag{58}$$

Let $C^2 = -\frac{\partial^2 \Phi}{\partial z^2}$. It may regarded as a kinematic parameter, not unlike A and B. The determination of C and D in the neighborhood of the Sun was attempted by Oort, Parenago, Kuzmin, Safronov. Probable values of these parameters are

$$C = 0.068 \text{ km/sec/pc},$$
$$D = 5.2 \cdot 10^{-24} \text{ g/cm}^3.$$

The density obtained from Eq. (58) is clearly the total density of stellar, diffuse, and other possible forms of matter in the system.

§159. **Variation of the residual velocity of a star. Circular and almost circular orbits.** We will now write the equations of motion of a star in cylindrical coordinates which are obtained from (30) assuming a system with circular axial symmetry and $z = 0$ as the plane of symmetry. From (30)

$$\left.\begin{aligned} \frac{d^2R}{dt^2} &= R\left(\frac{d\theta}{dt}\right)^2 + \frac{\partial \Phi}{\partial R}, \\ \frac{d}{dt}\left(R^2\frac{d\theta}{dt}\right) &= 0, \\ \frac{d^2z}{dt^2} &= \frac{\partial \Phi}{\partial z}. \end{aligned}\right\} \tag{59}$$

In the plane of symmetry $\frac{\partial \Phi}{\partial z}$ vanishes and the conditions

$$\left.\begin{aligned} &R = \text{const}, \\ &\frac{d\theta_c}{dt} = \sqrt{-\frac{1}{R}\frac{\partial \Phi}{\partial R}} = \text{const}, \\ &z = 0 \end{aligned}\right\} \tag{60}$$

are therefore clearly seen to satisfy Eqs. (59). Circular orbits satisfying these equalities are thus allowed in the stellar system.

Let us now consider the velocity of a star relative to a point which at a certain time had the same coordinates as a star moving in a circular orbit (60). Let ξ be the projection of the relative velocity on the plane of symmetry $z = 0$, α the angle this projection makes with the direction of the circular velocity, and η the projection of the relative velocity on the z axis. Then

$$\left.\begin{aligned} \frac{dR}{dt} &= \xi \sin\alpha, \\ R(\dot\theta - \dot\theta_c) &= \xi\cos\alpha, \\ \frac{dz}{dt} &= \eta. \end{aligned}\right\} \tag{61}$$

If we now write the first two equations in (59) in the form

$$\left.\begin{aligned} \frac{d^2R}{dt^2} &= R\dot\theta^2 - R\dot\theta_c^2 = R(\dot\theta - \dot\theta_c)(\dot\theta - \dot\theta_c + 2\dot\theta_c), \\ \frac{d}{dt}[R^2(\dot\theta - \dot\theta_c + \omega)] &= \frac{d}{dt}[R^2(\dot\theta - \dot\theta_c)] + \frac{d}{dR}(R^2\omega)\frac{dR}{dt} = 0, \end{aligned}\right\} \tag{62}$$

we get by (61)

$$\left.\begin{aligned} \frac{d}{dt}(\xi\sin\alpha) &= \frac{\xi^2}{R}\cos^2\alpha + 2\omega\xi\cos\alpha, \\ \frac{d}{dt}(R\xi\sin\alpha) &= 2BR\xi\sin\alpha, \end{aligned}\right\} \tag{63}$$

where ω is the angular velocity for a circular orbit and

$$B = -\frac{1}{2R}\frac{d}{dR}(R^2\omega).$$

If at this stage we restrict the discussion to stars which initially had $\eta = 0$, i.e., which move in the plane of symmetry, Eqs. (63) remain valid at any time.

It should be emphasized that the velocity of the star is not taken relative to some fixed body moving in a circular orbit. The velocity is determined instantaneously relative to an object whose circular orbit passes through the star's position at a given time. Therefore $\dot\theta_c$ and B in Eqs. (62) and (63) are variable.

Expanding the derivatives and solving Eqs. (63) for $\dot\xi$ and $\dot\alpha$, we get

$$\left.\begin{aligned} \dot\xi &= A\xi\sin 2\alpha, \\ \dot\alpha &= \frac{\xi}{R}\cos\alpha + 2A\cos^2\alpha - 2B, \end{aligned}\right\} \tag{64}$$

where

$$A = \omega + B.$$

If the initial velocity of a star relative to a point moving in a circular orbit is small, the star will move in a nearly circular orbit at least for some time, and the coefficients A and B will therefore vary very slowly.

Ignoring this variation, i.e., taking A and B to be constant, we readily derive an integral of Eqs. (64):

$$\xi^2 - \xi_1^2 + 4AR\xi\cos\alpha + 8ABR^2\ln\left(1 - \frac{\xi\cos\alpha}{2BR}\right) = 0, \tag{65}$$

where ξ_1 is the value of ξ for $\alpha = \frac{\pi}{2}$.

As we have noted in §157, the centroid of stars with small residual velocities moves (in a circular orbit) with a velocity which is nearly equal to circular. Therefore, if ξ and η are small, they can be considered as the projections of the star's residual velocity. In this case comparison with Eqs. (66), Ch. XIX shows that the coefficients A and B are the Oort coefficients.

If we take $\frac{\xi}{2BR} \ll 1$ (in the neighborhood of the Sun this ratio is approximately equal to 1/11), then ignoring terms of third and higher order in (65) we obtain

$$\xi^2 = \frac{B}{A\cos^2\alpha}\xi_1^2. \tag{66}$$

The same result can be obtained by integration of the equations

$$\left.\begin{aligned}\dot{\xi} &= A\xi\sin 2\alpha,\\ \dot{\alpha} &= 2A\cos^2\alpha - 2B,\end{aligned}\right\} \tag{67}$$

which are obtained from (64) when $\frac{\xi}{2BR}$ is small. Eq. (66) shows that for

$$\frac{A}{B} - 1 < 0 \tag{68}$$

the tip of the projection of the residual velocity vector on the galactic plane describes an ellipse with axes ξ_1 and $\xi_1\sqrt{\frac{B}{B-A}}$. Since $A - B = \omega > 0$, condition (68) is equivalent to the following:

$$B = \frac{1}{4}\sqrt{\frac{R}{\frac{\partial\Phi}{\partial R}}}\left[\frac{\partial^2\Phi}{\partial R^2} + \frac{3}{R}\frac{\partial\Phi}{\partial R}\right] < 0,$$

i.e.,

$$\frac{\partial^2\Phi}{\partial R^2} + \frac{3}{R}\frac{\partial\Phi}{\partial R} < 0. \tag{69}$$

In mechanics (69) is known as the stability condition of circular orbits. It is readily seen from (66) that if (68) is not satisfied, the residual velocity of the star increases indefinitely.

Let us now consider the distribution function of residual velocities, restricting the discussion as before to small velocities.

Each star has different velocities at different times and the set of these velocity values of a given star is described by some distribution function. The distribution of residual velocities of a star can be defined by two parameters: ξ_0, the projection of the residual velocity on the plane of

symmetry at the time when $\alpha = \frac{\pi}{2}$, and Z_0, the projection of the residual velocity on the symmetry axis at the time when the star crosses the plane of symmetry.

The time when $\alpha = \frac{\pi}{2}$ is clearly different for different stars; this also applies to the time of passage through the plane of symmetry. However, in a system which is stationary in the regular field the number of stars with $\alpha = \frac{\pi}{2}$ at any given time is independent of time and the number of stars passing through the plane of symmetry at a given time is similarly constant. Moreover, the number of stars with given ξ_0 and Z_0 which have $\alpha = \frac{\pi}{2}$ at a given time or cross the plane of symmetry at a given time is constant.

Therefore, the observed distribution of residual velocities is the distribution that we obtain by assuming an arbitrary law of passage of all the

$$f(\xi_0^2,\ Z_0^2)\, d\xi_0\, dZ_0$$

stars through states with $\alpha = \frac{\pi}{2}$ and $z = 0$ and then consider the entire set of residual velocities for all the times. A convenient assumption, say, is that at some time all the

$$f(\xi_0^2,\ Z_0^2)\, d\xi_0\, dZ_0 \tag{70}$$

stars have $\alpha = \frac{\pi}{2}$ and $z = 0$, from which we can proceed to determine the distribution function of their velocities at any time.

Consider the distribution function of residual velocities at any time for stars with given ξ_0. For the four-dimensional phase space (we are now concerned with the distribution of the projections of the residual velocities on the plane of symmetry) we can write an equation based on the fact that the number of points in a given volume element of the phase space is proportional to the time it takes the representing point to cross the volume element in the velocity space and the time it takes the star to cross a volume element in the ordinary space:

$$\psi(\xi^2,\ \alpha)\, dq_1\, dq_2\, \xi\, d\alpha\, d\xi = c(\xi_0^2)\,(dt)_1\,(dt)_2, \tag{71}$$

where q_1 and q_2 are some orthogonal coordinates of the star in a coordinate system lying in the plane of the stellar system, rotating with an angular velocity Ω, and having its origin at the center of the system, $(dt)_1$ is the time it takes the star to cross the volume element $dq_1\, dq_2$ in the geometrical space, $(dt)_2$ is the time it takes the representing point to cross the volume element $\xi\, d\alpha\, d\xi$ in the velocity space, $c(\xi_0^2)$ is proportional to the number of stars for which $\xi = \xi_0$ for $\alpha = \frac{\pi}{2}$.

We may regard dq_1, dq_2, $\xi d\alpha$, $d\xi$ as constant. Further note that since we are dealing with infinitesimal residual velocities, the displacements of the star are also infinitesimal and the phase density ψ is independent of the coordinates.

Let the coordinate line q_1 be tangent to the star's trajectory. Then

$$(dt)_1 = \frac{dq_1}{\xi}. \tag{72}$$

If $d\sigma$ is the arc element that the star traverses in the velocity space inside the volume $\xi\, da\, d\xi$, we get

$$(dt)_2 = \frac{d\sigma}{\sqrt{\xi^2\dot{\alpha}^2 + \dot{\xi}^2}} = \frac{1}{\dot{\xi}\dot{\alpha}}\xi\, da. \tag{73}$$

Equality (73) can be written in this form, since by (64) and the stability condition (68) $\dot{\alpha} > 0$.

Inserting (72) and (73) in (71) and seeing that dq_2 and $d\xi$ are constant, we get

$$\psi(\xi^2,\ \alpha) = c_1(\xi_0^2)\frac{1}{\xi^2\dot{\alpha}}, \tag{74}$$

where $c_1(\xi_0)$ is proportional to the number of stars with $\xi = \xi_0$ for $\alpha = \frac{\pi}{2}$.

Multiplying the second equality in (64) by ξ^2, we get

$$\xi^2\dot{\alpha} = 2(A\cos^2\alpha - B)\xi^2 = -2B\xi_0^2. \tag{75}$$

Eq. (74) can therefore be written in the form

$$\psi(\xi^2,\ \alpha) = c_1(\xi_0^2)\cdot\frac{1}{-2B\xi_0^2} = c_1\left(\frac{A\cos^2\alpha - B}{-B}\xi^2\right)\frac{1}{2(A\cos^2\alpha - B)\xi^2}. \tag{76}$$

Hence it follows that

$$\psi\left(\xi_0^2,\ \frac{\pi}{2}\right) = \psi\left[\left(\frac{A}{-B}\cos^2\alpha + 1\right)\xi^2,\ \alpha\right]. \tag{77}$$

We recall that ψ is the distribution function of the projections of the residual velocities on the plane of symmetry at any time, representative of all the stars having $\xi = \xi_0$ for $\alpha = \frac{\pi}{2}$. Since the distribution of stars over ξ_0 and Z_0 is expressed by (70), it follows from (77) and (70) that the distribution function of the residual velocities in the plane of symmetry of the stellar system is

$$f\left[\left(\frac{A}{-B}\cos^2\alpha + 1\right)\xi^2,\ Z_0^2\right].$$

Changing over to velocity components in the cylindrical system

$$\Pi = \xi\sin\alpha,$$
$$\vartheta - \vartheta_0 = \xi\cos\alpha,$$

we finally get

$$f\left[\Pi^2 + \frac{A-B}{-B}(\vartheta - \vartheta_0)^2,\ Z_0^2\right]. \tag{78}$$

Thus the distribution function of infinitesimal residual velocities in the plane of symmetry of a stationary stellar system with a symmetry axis

and a symmetry plane contains the components Π and $\vartheta-\vartheta_0$ not individually but in a certain combination $\Pi^2+\frac{A-B}{B}(\vartheta-\vartheta_0)^2$. In other words, these components follow an elliptic distribution.

Note that although the tip of the projection vector of the residual velocity on the plane of symmetry describes an ellipse (66), it does not follow that the velocity distribution is elliptic. To obtain an elliptic velocity distribution, the tip of the projection vector should move with a velocity consistent with the law of areas (75), which was actually used in passing from (74) to (76).

The function (78) also provides an approximate description of the distribution of finite residual velocities. The approximation improves as the residual velocities become smaller.

A measure of approximation is provided in particular by the ratio $\left|\frac{\zeta}{2B\Pi}\right|$, which was ignored compared to unity in passing from (65) to (66). For stars of the disk component of the Galaxy, this ratio is about 1/15, and for stars of the halo component it is close to 1/5.

Observations give the ratio $1/\sqrt{\frac{-B}{A-B}}$ for the ellipse semiaxes pointing to the center of the Galaxy and along the centroid velocity, respectively (see §150).

Note that the derived property of the distribution of residual velocities is entirely attributable to the regular field of the stellar system and is not a consequence of stellar encounters. Moreover, in the derivation we have tacitly assumed that the residual velocity varies only in the regular field and ignored the effect of stellar encounters.

The actual form of (78) clearly depends on the form of (70), which is unknown. It is also impossible to link up the velocity component ζ_0 with the components Π and $\vartheta-\vartheta_0$ by considering the regular field in general form. It is clear, however, that the ellipsoidal distribution does not apply, since otherwise observations would have given ratio (53), as well as (54), for the ellipsoid axes.

Let us determine the time for the projection of the residual velocity on the plane $z=0$ to complete one revolution. Integration of the second equation in (67) gives

$$t+c=\frac{1}{2\sqrt{-B(A-B)}}\operatorname{arctg}\left(\sqrt{\frac{-B}{A-B}}\operatorname{tg}\alpha\right).$$

As α goes from 0 to $\frac{\pi}{2}$, the right-hand side of the equality is incremented by $\frac{\pi}{4\sqrt{-B(A-B)}}$; the total period is thus

$$T=\frac{\pi}{\sqrt{-B(A-B)}}.$$

In general the period of complete revolution of the residual velocity vector is not commensurate with the galactic rotation period. Their ratio is

$$\frac{1}{2}\sqrt{\frac{A-B}{-B}},$$

which for the Galaxy is approximately equal to 0.8.

Focusing our attention on motion along the z axis, we note that for small z we approximately have

$$\frac{\partial \Phi}{\partial z} = \left(\frac{\partial^2 \Phi}{\partial z^2}\right)_{z=0} \cdot z = -C^2 z.$$

The third equation in (59) therefore has a solution

$$z = a \sin C(t - t_1), \tag{79}$$

where t is the time of passage through the plane $z = 0$, and a is the maximum distance of the star from this plane. Differentiation of (79) with respect to t gives

$$\eta = Ca \cos C(t - t_1).$$

The period of η is $\frac{2\pi}{C}$.

As we have alrady noted, C is approximately equal to 0.068 km/sec/pc = $= 2.2 \times 10^{-15}$ sec^{-1}, so that the period of η, i.e., the period of oscillation of a star about the plane of symmetry, is about $9 \cdot 10^7$ years.

§160. The virial theorem and its consequences. Consider a system which is stationary in relation to regular forces. The parameters of this system vary only due to the action of irregular forces, and their variation is therefore slow. In this case the variation of the system parameters over fairly large time periods is adequately represented by a linear function. In particular, writing the expression for the moment of inertia of a stellar system relative to a fixed origin,

$$I = m \sum_i (x_i^2 + y_i^2 + z_i^2). \tag{80}$$

where the sum is taken over all the stars in the system (equal masses are assumed for all the stars), we may take

$$\frac{d^2 I}{dt^2} = 0. \tag{81}$$

Insertion of (80) in (81) gives

$$\frac{1}{2}\frac{d^2 I}{dt^2} = m \sum_i (\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2) + m \sum_i (x_i \ddot{x}_i + y_i \ddot{y}_i + z_i \ddot{z}_i) = 0. \tag{82}$$

The first sum in this expression is double the kinetic energy of the stars in the system:

$$2T = m \sum_i w_i^2, \tag{83}$$

where w_i is the overall velocity of a star.

Let U be the potential energy of the stellar system, which is equal to the gravitational interaction energy of all the stars in the system. Thus,

$$U = -Gm^2 \sum_{i \neq j} \sum \frac{1}{r_{ij}}, \tag{84}$$

where a double sum is used, since all the pair combinations of the stars are considered. The equations of motion for each star have the form

$$m\ddot{x} = -\frac{\partial U}{\partial x_i}; \quad m\ddot{y} = -\frac{\partial U}{\partial y_i}; \quad m\ddot{z} = -\frac{\partial U}{\partial z_i}.$$

The second sum in (82), called the virial of the system, can be transformed as follows:

$$m\sum_i (x_i\ddot{x}_i + y_i\ddot{y}_i + z_i\ddot{z}_i) = -\sum_i \left(x_i\frac{\partial U_i}{\partial x_i} + y_i\frac{\partial U}{\partial y_i} + z_i\frac{\partial U}{\partial z_i}\right). \tag{85}$$

The potential energy of the system according to (84) is a homogeneous function of degree -1 of the coordinates x_i, y_i, z_i, and therefore by Euler's theorem the right-hand side of (84) is simply equal to U.

Eq. (82) thus takes the form

$$2T + U = 0, \tag{86}$$

i.e., in a stellar system which is stationary relative to regular forces the kinetic energy is equal to minus half the potential energy. This is the virial theorem.

Since the total energy of a stellar system is the sum of kinetic plus potential energy,

$$E = T + U,$$

it follows from (86) that

$$E = -T = \tfrac{1}{2}U, \tag{87}$$

i.e., the total energy of a system which is stationary relative to regular forces is equal to half the potential energy or to minus the kinetic energy.

Let us find an approximate value of the potential energy for a stellar system which is a homogeneous oblate ellipsoid of revolution. Note that the number of terms in the right-hand side of (84) is $\frac{1}{2}N(N-1) \approx \frac{1}{2}N^2$, where N is the number of stars in the system. This sum is thus equal to $\frac{1}{2}N^2$ times the harmonic mean of all the pair distances between the points of the ellipsoid,

$$\left(\frac{1}{r}\right) = \frac{1}{\delta^2}\int\int \frac{1}{r}\,d\gamma_1\,d\gamma_2. \tag{88}$$

where δ is the volume of the ellipsoid, $d\gamma_1$ and $d\gamma_2$ are the ellipsoid volume elements, and the integration is carried out twice over the entire ellipsoid. The integral in (88) cannot be evaluated analytically, but numerical integration gives

$$\left(\frac{1}{r}\right) = \frac{1}{B\varrho}, \tag{89}$$

where ϱ is the major radius of the ellipsoid, and B is a parameter which depends on the oblateness of the ellipsoid; it decreases with increasing

oblateness and satisfies the inequality

$$0.53 \approx \frac{5}{3\pi} = B_1 \ll B \ll B_2 = \frac{5}{6} \approx 0.83, \tag{90}$$

where B_1 corresponds to perfectly flat (oblateness $h = 1$) and B_2 to spherical ($h = 0$) system, assuming that these systems are homogeneous.

Thus,

$$U = -\frac{Gm^2N^2}{2B\varrho} \tag{91}$$

and by (87) the total energy of the system is

$$E = -\frac{Gm^2N^2}{4B\varrho}. \tag{92}$$

Eq. (83) can be written in the form

$$T = \frac{1}{2} Nm\bar{w}^2, \tag{93}$$

where w is the rms velocity of stars in the system.

Inserting (91) and (93) in (87) and solving for $\bar{w}$, we get

$$\bar{w} = \sqrt{\frac{GmN}{2B\varrho}}. \tag{94}$$

Eq. (94) expresses the rms velocity of stars in a system in terms of other parameters; for the Galaxy, taking $mN = 10^{44}$ g, $\rho = 10^4$ pc, $B = 0.7$, we get $\bar{w} \cong 120$ km/sec. For a globular cluster ($mN = 10^{39}$g, $\rho = 20$ pc, $B = 0.8$), $\bar{w} \cong 7$ km/sec. For an open cluster ($mN = 2 \cdot 10^{35}$g, $\rho = 2$ pc, $B = 0.8$), $\bar{w} \cong 0.5$ km/sec.

Let us now find the escape energy of a star, i.e., the energy required to overcome the gravitational attraction of the system. The escape energy is clearly equal to minus the potential energy U' of the star in the system

$$-U'_j = Gm^2 \sum_{i \neq j} \frac{1}{r_{ij}}, \tag{95}$$

where the subscript j indicates that this is the escape energy of the j-th star, and the single sum is taken over all i, except $i = j$.

Comparison of (84) and (95) shows that the number of terms in (84) is $\frac{1}{2}N(N-1)$, and in (95) $N-1$. Therefore if we formally add up the potential energies of the individual stars in the system, the result is double the potential energy of the system as a whole:

$$\sum_j U'_j = 2U = -\frac{Gm^2N^2}{B\varrho}. \tag{96}$$

Dividing (96) through by N, we obtain the average potential energy of a star in the system. Minus this energy,

$$\frac{Gm^2N}{B\varrho}, \tag{97}$$

is the mean escape energy of a star. This energy is equal to double the potential energy of the system per star. This relation also emerges from the following considerations. If we were to detach all the stars one by one from the system, the total energy expended in the process would be equal (in magnitude) to the potential energy of the system. However, as the removal of each star from the system reduces the total stellar population, the escape energy of each successive star will be less than that of its percussor, since it is proportional to the number of remaining stars minus one. Hence it follows that the escape energy of the first star to be removed from the system, (97), is double the potential energy of the system per star. It also follows from (96) that writing the formal sum of energies of all the individual stars we get

$$\sum_i E_i = \sum_i (U_i + T_i) = 2U + T. \tag{98}$$

since the kinetic energy of the system is equal to the sum of the kinetic energies of the constituent stars.

Now consider a star whose kinetic energy is equal to the mean escape energy (for the first star). If the velocity of this star, which is the critical velocity, is designated ω_k, we have

$$\frac{1}{2} m\omega_k^2 = \frac{Gm^2N}{BQ}.$$

Therefore

$$\omega_k = \sqrt{\frac{2GmN}{BQ}}. \tag{99}$$

Comparison of (99) and (94) gives

$$\omega_k = 2\bar{\omega}, \tag{100}$$

i.e., the mean critical velocity of a star is equal to double the rms velocity.

§161. **Irregular forces. Variation of the absolute velocities of stars following binary encounters.** Let us first consider the relative velocities. We write w for the velocity vector of the star S relative to the star S_1 before the encounter and place the origin at the star S_1 (Figure 180).

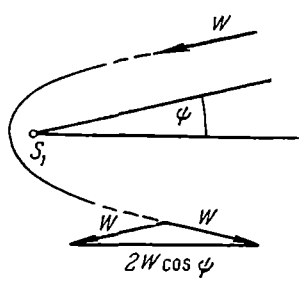


FIGURE 180

Let m and m_1 be the masses of the two stars, respectively. As we know, the star S moves in a hyperbola whose equation in polar coordinates is

$$r = \frac{\sigma}{1 - e\cos\varepsilon}. \tag{101}$$

if the polar axis is directed along the axis of the hyperbola. Here e is the eccentricity, σ is the parameter of the hyperbola. Let ψ be the angle between the axis of the hyperbola and its asymptote. Then, as we see from the figure, the direction of the velocity vector of the star S will change by $\pi - 2\psi$ in the encounter. According to the law of energy conservation, the magnitude of the relative velocity does not change in such an encounter, and the magnitude of the vector increment of the relative velocity is

$$2\omega\cos\psi. \tag{102}$$

Let us find $\cos\psi$. We see from (101) that $\varepsilon \to \psi$ for $r \to \infty$. Thus,

$$\cos\psi = \frac{1}{e} \tag{103}$$

We know from analytical geometry that

$$\frac{\sigma^2}{p^2} = e^2 - 1. \tag{104}$$

where p is the impact parameter. On the other hand, from the law of areas in the two-body problem

$$\sqrt{G(m + m_1)} = \omega p, \tag{105}$$

since the right-hand side of (105) is double the areal velocity at infinity.

Eliminating e and σ between (103), (104), and (105), we get

$$\cos\psi = \left[1 + \frac{p^2\omega^4}{G^2(m + m_1)^2}\right]^{-\frac{1}{2}}. \tag{106}$$

The change in relative velocity is thus obtained without much difficulty. Multiplying the result by $\frac{2m}{m+m_1}$, we obtain the change in velocity relative to the center of inertia of the two approaching stars.

We are naturally concerned with the change in the absolute velocities, i.e., the velocities of stars in some fixed frame of reference.

If v and v_1 are the velocities of the stars S and S_1 before the encounter in a fixed frame of reference, we have

$$w = v - v_1, \tag{107}$$

$$v_g = \frac{1}{m + m_1}(mv + m_1 v_1), \tag{108}$$

where v_g is the velocity of the center of inertia of the two stars, which remains constant during the encounter.

From (107) and (108) we readily find

$$v = v_g + \frac{m_1}{m + m_1} w.$$

whence

$$v^2 = v_g^2 + \frac{2m_1}{m+m_1} v_g w \cos\chi + \frac{m_1^2}{(m+m_1)^2} w^2, \tag{109}$$

where χ is the angle between v_g and w. After the encounter, we similarly have

$$v'^2 = v_g^2 + \frac{2m_1}{m+m_1} v_g w \cos\chi' + \frac{m_1^2}{(m+m_1)^2} w^2, \tag{110}$$

where χ is the angle between v_g and the relative velocity w' at the end of the encounter.

Subtracting (109) from (110), we find the change in the square of the absolute velocity:

$$\Delta v^2 = \frac{2m_1}{m+m_1} v_g w(\cos\chi' - \cos\chi). \tag{111}$$

The plane through v and v_1 we call the fundamental plane. The vectors v_g and w also lie in this plane. The orbit in general does not lie in the fundamental plane. Let θ be the angle between the orbital plane and the fundamental plane. The vectors w and w' and hence the angle ψ all lie in the orbital plane.

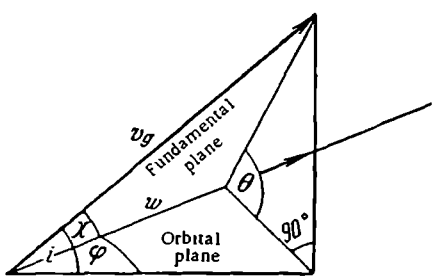


FIGURE 181

Let i be the angle between v_g and the orbital plane, φ and φ' the angles that w and w' make with the projection of v_g on the orbital plane. Then from Figure 181

$$\left.\begin{aligned} \cos\chi &= \cos\varphi\cos i;\\ \cos\chi' &= \cos\varphi'\cos i;\\ \sin\chi\cos\theta &= \sin\varphi\cos i. \end{aligned}\right\} \tag{112}$$

Eq. (111) can therefore be written in the form

$$\Delta v^2 = \frac{2m_1}{m+m_1} v_g w(\cos\varphi' - \cos\varphi)\cos i. \tag{113}$$

The angles φ and φ' lie in the orbital plane and are related to the angle ψ by the equality

$$\varphi' - \varphi = \pi - 2\psi. \tag{114}$$

Using in (113) the trigonometric identity for the difference of two cosines and applying (114), we get

$$\Delta v^2 = -\frac{4m_1}{m+m_1} v_\varrho \omega \cos(\varphi - \psi)\cos\psi\cos\iota. \tag{115}$$

Expanding the cosine of the difference of two angles and using (106) and (112), we get

$$\Delta v^2 = -\frac{4m_1}{m+m_1}\frac{v_\varrho\omega}{1+\frac{p^2\omega^4}{G^2(m+m_1)^2}}\left[\cos\chi + \sin\chi\cos 0\,\frac{p\omega^2}{G(m+m_1)}\right]. \tag{116}$$

Further using the properties of the scalar and vector products, we find

$$\omega v_\varrho\cos\chi = \frac{1}{m+m_1}(\boldsymbol{v} - \boldsymbol{v}_1 m\boldsymbol{v} + m_1\boldsymbol{v}_1) =$$
$$= \frac{1}{m+m_1}[mv^2 - m_1v_1^2 + (m_1 - m)vv_1\cos\alpha], \tag{117}$$

$$\omega v_\varrho\sin\chi = \frac{1}{m+m_1}[\boldsymbol{v} - \boldsymbol{v}_1 \times m\boldsymbol{v} + m_1\boldsymbol{v}_1] = vv_1\sin\alpha. \tag{118}$$

Making use of (117) and (118) in (116), we finally obtain for the change in the square of the absolute velocity of a star following an encounter

$$\Delta v^2 = -\frac{4m_1}{(m+m_1)^2}\frac{1}{1+\frac{p^2\omega^4}{G^2(m+m_1)^2}}\times$$
$$\times\left[mv^2 - m_1v_1^2 + (m - m_1)vv_1\cos\alpha + \frac{p\omega^2}{G}vv_1\sin\alpha\cos\theta\right]. \tag{119}$$

Eq. (119) was originally derived by Chandrasekhar.

§162. The probability of a binary encounter with a given change in absolute velocity. Let $\beta = \frac{v^2}{\bar{v}^2}$, where $\bar{v}$ is the rms velocity. We will find the probability $F(\beta, h)dh\,dt$ that in a time dt a star with a given β will encounter another field star under such conditions that the relative change in the square of its velocity $h = \frac{\Delta\beta}{\beta} = \frac{\Delta v^2}{v^2}$ will be between h and $h + dh$. We will only consider binary encounters, ignoring encounters between more than two stars. The problem can be investigated for a general case of an arbitrary mass distribution function in the system (which is however assumed to be the same throughout the system).

Here for simplicity we will consider two important particular cases:

A. All the field stars have the same mass, which is further equal to the mass of the star being considered.

B. All the field stars have the same mass, but the mass of the star being considered is vanishingly small.

Let us first consider case A. Eq. (119) can now be written in the form

$$\cos\theta = -\frac{\mu}{p\bar{h}\omega^2\sin\alpha}\left[\left(1 + \frac{p^2\omega^4}{4\mu^2}\right)h + 1 - k^2\right], \tag{120}$$

where

$$k = \frac{v_1}{v}, \quad \mu = Gm. \tag{121}$$

For the variables in (120) we have the constraints

$$-1 \leqslant h \leqslant k^2, \tag{122}$$

$$k \geqslant 0, \tag{123}$$

$$p \geqslant 0, \tag{124}$$

$$-1 \leqslant \cos\alpha \leqslant +1, \tag{125}$$

$$1 - \cos^2\theta \geqslant 0. \tag{126}$$

Constraint (122) shows that the square of the velocity cannot decrease by more than the square of the velocity of the star itself and cannot increase by more than the square of the velocity of the field star taking part in the encounter (the latter follows from energy conservation considerations).

Let now all the parameters, except $\cos\theta$ and h, be fixed. Then

$$d\cos\theta = -\frac{\mu}{p\omega^2 k \sin\alpha}\left(1 + \frac{p^2\omega^4}{4\mu^2}\right) dh. \tag{127}$$

If the star field is characterized by a star density D, and the velocity distribution in the field is described by $f(k)\,dk$, the number of encounters between the star being considered and the field stars in a time dt for the given parameters is

$$2\pi p\,dp \cdot D\omega\,dt \cdot f(k)\,dk \cdot \frac{1}{2}\sin\alpha\,d\alpha \cdot \frac{d\cos\theta}{\pi\sqrt{1-\cos^2\theta}}. \tag{128}$$

Consider the four-dimensional space (k, $\cos\alpha$, p, $\cos\theta$) and its subspace (k, $\cos\alpha$, p). We fix the values of h and $h + dh$. If $\cos\theta$ in inequality (126) is replaced by its expression from (120), the set of inequalities (122) through (126) define a domain in the subspace (k, $\cos\alpha$, p) which ensures the desired h. Each point in this part of the subspace corresponds in the space (k, $\cos\alpha$, p, $\cos\theta$) to a point which generates the given h and another point which generates the given $h + dh$. The values of the coordinate $\cos\theta$ in this subspace are defined by inequalities (120) and (127). In this way we obtain the set of points in the space (k, $\cos\alpha$, p, $\cos\theta$) which give h between h and $h + dh$. The probability density in this region is given by (128).

It follows from the preceding that to find the probability of an encounter with h lying between h and $h + dh$, we should insert for $\cos\theta$ and $d\cos\theta$ in (128) their expressions from (120) and (127) and then integrate over the entire domain of the subspace (k, $\cos\alpha$, p) defined by constraints (122)—(126); in (126), $\cos\theta$ should also be replaced by its expression from (120).

The domain of integration over p is determined by constraints (124) and (126). Inserting (120) in (126) and writing

$$a = \frac{\omega^2 h}{4k\mu\sin\alpha}, \tag{129}$$

$$b = \frac{\mu}{\omega^2 k\sin\alpha}(h - 1 - k^2), \tag{130}$$

$$c = \frac{\mu}{\omega^2 k\sin\alpha}, \tag{131}$$

we obtain

$$a^2p^4 - (1 - 2ab)\,p^2 + b^2 < 0. \tag{132}$$

This inequality is solvable if

$$(1 - 2ab)^2 - 4a^2b^2 = 1 - 4ab > 0. \tag{133}$$

If, however, (133) is satisfied, the two roots $(p^2)_1$ and $(p^2)_2$ of (132) are positive. Writing $p_1 = \sqrt{(p^2)_1}$ and $p_2 = \sqrt{(p^2)_2}$, we see that for positive p inequality (133) is satisfied when

$$p_1 \leqslant p \leqslant p_2, \tag{134}$$

which defines the range of integration.

Inserting from (120) and (127) in (128), we carry out the integration over p and using the notation (129)—(131) write

$$D w\, dt\, f(k)\, dk \sin\alpha\, d\alpha \int_{p_1}^{p_2} \frac{\left(c + \frac{a}{h} p^2\right) p\, dp}{\sqrt{-|a^2 p^4 - (1-2ab)\, p^2 + b^2}}. \tag{135}$$

Changing over to p^2 as the variable of integration, using Euler's substitution, and seeing that $(p^2)_1$ and $(p^2)_2$ are the roots of the radicand, we obtain after integration making use of (129)—(131)

$$2\pi D \mu^2\, dt \frac{dh}{|h^3|} f(k)\, dk \left[2k^2 \sin^2\alpha - h(1-k^2) \right] \frac{d(\cos\alpha)}{w^3}. \tag{136}$$

We can now integrate (136) over $\cos\alpha$ in a certain interval $[(\cos\alpha)_1,\ (\cos\alpha)_2]$. To this end, we make use of the relation between the absolute and the relative velocities of the approaching stars:

$$w^2 = v^2 + v_1^2 - 2vv_1 \cos\alpha = v^2(1 + k^2 - 2k\cos\alpha). \tag{137}$$

Changing over to a new variable of integration k and taking the integral, we get

$$\frac{\pi}{v^3} \frac{dh}{|h^3|} D\, dt\, \mu^2 f(k)\, dk \cdot \frac{1}{k} \left[-\frac{v}{w}(1-k^2)(1+2h-k^2) - 2\frac{w}{v}(1+k^2) + \frac{w^3}{3v^3} \right]_{w_1}^{w_2} \tag{138}$$

We insert (129) and (130) in (133). Factoring, we get

$$\cos^2\alpha \leqslant (1+h)\left(1 - \frac{h}{k^2}\right). \tag{139}$$

This inequality and inequality (125) specify the range of integration over $\cos\alpha$.

If

$$\left.\begin{aligned} &h > 0, \\ &h \leqslant k^2 < 1 + h, \end{aligned}\right\} \tag{140}$$

we have

$$(1+h)\left(1 - \frac{h}{k^2}\right) - 1 = \frac{h}{k^2}(k^2 - 1 - h) \leqslant 0,$$

so that the range of integration over $\cos\alpha$ is defined by inequality (139). In this case we calculate w_1 and w_2:

$$w_{1,2} = \sqrt{v^2 \left\{ 1 + k^2 - 2k \left[\mp \sqrt{(1+h)\left(1 - \frac{h}{k^2}\right)} \right] \right\}} = v\left(\sqrt{1+h} \pm \sqrt{k^2 - h}\right). \tag{141}$$

Inserting these values of ω_1 and ω_2 in (138), we find after simple manipulations

$$\frac{\pi}{v^3}\frac{dh}{|h^3|}D\mu^2 f(k)\,dk\,dt\,\frac{1}{k}\,\frac{4}{3}(4k^2-h)\sqrt{k^2-h}. \tag{142}$$

If now

$$\left.\begin{aligned} h &> 0, \\ k^2 &\gg 1+h, \end{aligned}\right\} \tag{143}$$

the limits of integration over cos α are defined by inequality (125), and ω_1 and ω_2 are expressed by

$$\omega_{1,2} = \sqrt{v^2[1+k^2-2k(+1)]} = v(k \pm 1). \tag{144}$$

Insertion of ω_1 and ω_2 in (138) gives

$$\frac{\pi}{v^3}\frac{dh}{|h^3|}D\mu f(k)\,dk\,dt\,\frac{1}{k}\,4\left(h+\frac{4}{3}\right). \tag{145}$$

Using (140) and (143), we can now take the integral over k, which for $h>0$ gives the sought probability of encounter with a specified change in velocity:

$$F(\beta, h)\,dh\,dt = \frac{\pi}{v^3\beta^{3/2}}\frac{dh}{h^3}D\mu^3\,dt\left[\int_{\sqrt{h}}^{\sqrt{1+h}} f(k)\,dk\cdot\frac{1}{k}\,\frac{4}{3}(4k^2-h)\sqrt{k^2-h}\,dk + \right.$$
$$\left. + \int_{\sqrt{1+h}}^{\infty} f(k)\,dk\cdot\frac{1}{k}\,4\left(h+\frac{4}{3}\right)dk\right]. \tag{146}$$

Assuming a Maxwellian distribution,

$$f(k)\,dk = f(v_1)\,dv_1 = \frac{3\sqrt{6}}{\sqrt{\pi}}\frac{v_1^2}{v^3}e^{-\frac{3}{2}\frac{v_1^2}{v^2}}dv_1 = \frac{3\sqrt{6}}{\sqrt{\pi}}\beta^{\frac{3}{2}}k^2e^{-\frac{3}{2}\beta k^2}dk. \tag{147}$$

Inserting (147) in (146), changing over to a new variable of integration $k_2=\sqrt{k^2-h}$, integrating in the second term and integrating by parts in the first term, we obtain after simplifications

$$F(\beta, h) = \frac{4\sqrt{6\pi}\,D\mu^2}{v^3\beta h^3}\int_0^1 (4k^2+h)\,e^{-\frac{3}{2}\beta(k^2+h)}\,dk. \tag{148}$$

To find the sought probability for $h<0$, note that in this case

$$k^2 \ll 1+h \tag{149}$$

leads to

$$(1+h)\left(1-\frac{h}{k^2}\right)-1 \geqslant 0. \tag{150}$$

This shows that the integral should be taken between the limits specified by inequality (125) and $\omega_{1,2}=v(1\pm k)$. Substitution of ω_1 and ω_2 in (138) thus gives

$$\frac{\pi}{v^3}\frac{dh}{-h^3}D\mu f(k)\,dk\,dt\,dh\cdot 4\left(\frac{4}{3}k^2-h\right). \tag{151}$$

Similarly, the conditions

$$\left.\begin{array}{l} h<0, \\ 1+h<k^2 \end{array}\right\} \tag{152}$$

give

$$\omega_{1,2}=v\left(\sqrt{k^2-h}\pm\sqrt{1+h}\right). \tag{153}$$

and from (138) we have

$$\frac{\pi}{v^3}\frac{dh}{-h^3}D\mu f(k)\,dk\cdot\frac{1}{k}\frac{4}{3}(4+h)\sqrt{1+h}\,. \tag{154}$$

The integral over k takes the form

$$\frac{\pi}{v^3}\frac{dh}{-h^3}d\mu\,dt\left[\int_0^{\sqrt{1+h}}4\left(\frac{4}{3}k^2-h\right)f(k)\,dk+\int_{\sqrt{1+h}}^{\infty}\frac{4}{3}(4+h)\sqrt{1+h}\,\frac{1}{k}f(k)\,dk\right] \tag{155}$$

Inserting (147) in (155) and transforming as for (146), we get

$$F(\beta,\ h)=\frac{4\sqrt{6\pi}\,D\mu^2}{-h^3\beta v^3}\int_0^{\sqrt{1+h}}(4k^2-h)\,e^{-\frac{3}{2}\beta k^2}\,dk. \tag{156}$$

It is now more convenient to replace $h=\frac{\Delta\beta}{\beta}$ by $\Delta\beta$, which is henceforth designated by g, so that $g=\Delta v^2-\beta h$. Now, seeing that

$$\Phi(\beta,\ g)\,dg\,dt=F(\beta,\ h)\,dh\,dt.$$

where $\Phi(\beta,\ g)\,dg\,dt$ is the probability of a stellar encounter changing β by an amount g occurring in a time dt, Eqs. (148) and (156) can be written in the form

$$\Phi(\beta,\ g)=\frac{\mu^2 D}{v^3}\Phi^*(\beta,\ g), \tag{157}$$

where

$$\Phi^*(\beta,\ g)=16\sqrt{6\pi}\,\frac{\beta}{|g^3|}\int_0^{\sqrt{1+\frac{|g|-g}{2\beta}}}\left(k^2+\frac{|g|}{4\beta}\right)e^{-\frac{3}{2}\left(\beta k^2+\frac{|g|+g}{2}\right)}\,dk. \tag{158}$$

Using the absolute value notation, we combined Eqs. (148) and (156) into a single expression, which applies to positive and negative g.

If for case B we take $m=0$, then acting along the same lines as in case A, we end up with the same expression (157), where $\Phi^*(\beta,\ g)$ is given by

$$\Phi^*(\beta,\ g)=16\sqrt{6\pi}\,\frac{\beta}{|g^3|}\int_{\frac{1}{2}\left|\sqrt{1+\frac{g}{\beta}}-1\right|}^{\frac{1}{2}\left(\sqrt{1+\frac{g}{\beta}}+1\right)}k^2e^{-\frac{3}{2}\beta k^2}\,dk. \tag{159}$$

The function $\Phi^*(\beta,\ g)$ is independent of the star field characteristics μ, D, $\bar{v}$. The values of $\lg\Phi^*(\beta,\ g)$ for case A, $m=m_1$, are tabulated in Table 1, and the values of $\lg\Phi^*(\beta,\ g)$ for case B, $m=0$, are tabulated in Table 2.

TABLE 1

g \ β	0	1.0	2.0	4.0
1.0	0.588	0.685	0.634	0.513
0.6	1.292	1.508	1.473	1.357
0.4	1.774	2.104	2.070	1.968
0.2	2.507	3.063	3.056	2.948
0.1	3.175	3.990	3.992	3.886
0.05	3.809	4.902	4.910	4.806
−0.05		4.918	4.939	4.839
−0.1		4.024	4.050	3.951
−0.2		3.126	3.166	3.077
−0.4		2.218	2.310	2.227
−0.6		1.648	1.815	1.745
−1.0		∞	1.186	1.148

TABLE 2

g \ β	0	1.0	2.0	4.0
1.0	1.077	1.106	1.045	0.921
0.6	1.586	1.745	1.707	1.587
0.4	1.978	2.256	2.230	2.114
0.2	2.605	3.135	3.128	3.017
0.1	3.223	4.023	4.027	3.920
0.05	3.834	4.921	4.928	4.823
−0.05		4.904	4.924	4.826
−0.1		3.993	4.019	3.922
−0.2		3.070	3.111	3.017
−0.4		2.112	2.198	2.111
−0.6		1.512	1.657	1.578
−1.0		∞	0.976	0.904

Eqs. (158) and (159) show that for $g \to 0$ the functions $\Phi^*(\beta, g)$ increase to infinity. This is so because in our integrals we did not impose an upper bound value on the impact parameter. However, it is only for distances less than the average interstellar distance that binary encounters are indeed much more frequent than multiple encounters, which were ignored in our analysis. Multiple encounters obviously reduce the overall effect, since the different stars pull in different directions and the geometrical sum of the gravitational forces is less than their arithmetic sum.

The complete solution should of course take multiple encounters into consideration. This complex problem, however, falls beyond the scope of our course. We will therefore consider an approximate solution. Let g_0 be the maximum change in β, when for both stars it is equal to unity, and the impact parameter is equal to the mean distance between stellar neighbors. Since for the average values of m and $\bar{\omega}^2$ in stellar systems and for p_0 of the order of 1 pc, $\frac{p\bar{\omega}^2}{mG}$ is of the order of 10^5, the last term in brackets in Eq.(119) is large compared to the other terms, which are therefore ignored, together with the unity in the denominator. We thus get

$$g_0 = \frac{\Delta v^2}{\bar{v}^2} \approx \frac{4\mu}{p_0\bar{\omega}^2} \approx \frac{2Gm}{p_0\bar{v}^2}. \qquad (160)$$

For $|g| > g_0$, the functions $\Phi^*(\beta, g)$ are given by (158) and (159). For $|g| < g_0$, $\Phi^*(\beta, g)$ do not vanish, since even for small impact parameters the velocity will change given an appropriate choice of the remaining encounter parameters. Nevertheless, the values of these functions are substantially less than those prescribed by Eqs. (158) and (159). The importance of these small velocity changes is not large, and the solution of the irregular field problem will not be distorted if for $|g| < g_0$ we take $\Phi^*(\beta, g) = \Phi^*(\beta, g_0)$.

Table 3 lists the values of g_0 for the star field around the Sun and for a globular and an open cluster.

TABLE 3

Stellar system	m, g	p_0, pc	$\bar{v}$, km/sec	g_0
Star field near the Sun	10^{33}	1.5	$3 \cdot 10^6$	$3 \cdot 10^{-6}$
Globular cluster	$2 \cdot 10^{33}$	0.4	$7 \cdot 10^5$	$5 \cdot 10^{-4}$
Open cluster	$2 \cdot 10^{33}$	0.6	$5 \cdot 10^4$	0.05

§163. The general problem of velocity variation in the irregular field. Relaxation time. If for some random variable β we know the probability $\Phi(\beta, g)\, dg\, dt$ that in a time dt the change in β will fall between g and $g + dg$ and the probability of noticeable changes is not particularly small compared to the probability of infinitesimal changes, we are dealing with a stationary discrete process of variation of the random variable β. Discrete processes are the subject of the theory of stochastic processes. The behavior of a random variable in a stationary discrete process is completely determined if we know the probability $L(\beta, t, \gamma)$ that a random variable with initial value β will assume after a time t some value not exceeding γ. The functions $\Phi(\beta, g)$ and $L(\beta, t, \gamma)$ satisfy the two integro-differential Kolmogorov—Feller equations:*

$$\left.\begin{aligned} \frac{\partial L(\beta, t, \gamma)}{\partial t} &= -L(\beta, t, \gamma)\int_0^\infty \Phi(\beta, \delta-\beta)\, d\delta + \int_0^\infty L(\delta, t, \gamma)\,\Phi(\beta, \delta-\beta)\, d\delta, \\ \frac{\partial L(\beta, t, \gamma)}{\partial t} &= \int_{-\infty}^{\gamma} L'_\delta(\beta, t, \delta)\, d\delta \int_0^\infty \Phi(\beta, \varepsilon-\delta)\, d\varepsilon - \int_0^\infty L'_\delta(\beta, t, \delta)\, d\delta \int_0^\gamma \Phi(\delta, \varepsilon-\delta)\, d\varepsilon. \end{aligned}\right\} \quad (161)$$

Since the function $\Phi(\beta, g)$ is known, Eqs. (161) can be used to find $L(\beta, t, \gamma)$.

Knowledge of $L(\beta, t, \gamma)$ enables us, in particular, to estimate the relaxation time of the system. Indeed, if $f(\beta)$ is the distribution function of the velocities squared (divided by $\bar{v}^2$) in a system which is stationary in an irregular field, we have

$$\lim_{t\to\infty} L(\beta, t, \gamma) = \int_0^\gamma f(\delta)\, d\delta,$$

$$\lim_{t\to 0} L(\beta, t, \gamma) = \begin{cases} 1, & \text{if } \beta < \gamma, \\ 0, & \text{if } \beta > \gamma. \end{cases}$$

The relaxation time τ can be defined as the time in which the probability L changes by 0.9 of the maximum change attainable for $t \to \infty$:

$$\frac{L(\beta, \tau, \gamma) - L(\beta, 0, \gamma)}{L(\beta, \infty, \gamma) - L(\beta, 0, \gamma)} = 0.9. \quad (162)$$

Although the value of τ in (162) depends on β and γ, this dependence is not very pronounced. We may therefore take the value of τ for $\beta = 0$, $\gamma = 1$, say.. Then, assuming a Maxwellian velocity distribution

$$f(\beta)\, d\beta = \frac{3\sqrt{3}}{\sqrt{2\pi}} \sqrt{\beta}\, e^{-\frac{3}{2}\beta}\, d\beta \quad (163)$$

we have

$$L(0, \tau, 1) = 0.1 + 0.9\int_0^1 f(\delta)\, d\delta = 0.524 \quad (164)$$

and τ can be found from (164).

The solution of Eqs. (161), however, involves considerable difficulties. We therefore recommend an alternative method for estimating relaxation times.

* See, e.g., Gnedenko, B. V. Kurs teorii veroyatnostei (A Course in Probability Theory), § 53.— Gostekhizdat. 1954.

The mean square rate of change of β for an average star is

$$\varkappa = \int_0^\infty f(\beta)\, d\beta \left[\int_0^\infty \Phi(\beta,\ g)\, g^2\, dg + \int_0^\beta \Phi(\beta,\ -g)\, g^2\, dg \right]. \qquad (165)$$

Inserting (157) in (165) and assuming a Maxwellian velocity distribution, we find

$$\varkappa = K \frac{\mu^2 D}{\bar{v}^3}, \qquad (166)$$

where K is a constant number:

$$K = \frac{3\sqrt{3}}{\sqrt{2\pi}} \int_0^\infty \sqrt{\beta}\, e^{-\frac{3}{2}\beta}\, d\beta \left[\int_0^\infty \Phi^*(\beta,\ g)\, g^2\, dg + \int_0^\beta \Phi^*(\beta,\ -g)\, g^2\, dg \right]. \qquad (167)$$

In accordance with the remarks at the end of §162, $\Phi^*(\beta, \pm g)$ can be taken equal to $\Phi^*(\beta, \pm g_0)$ when integrating over g from 0 to g_0. Evaluation of (167) for the values of g_0 listed in Table 3 gives the following values of K: around the Sun in the Galaxy 210, for a globular cluster 130, for an open cluster 50.

The relaxation time can be defined as the square of the range of the change in β, corresponding to a change in velocity by one rms velocity (this quantity is equal to unity), divided by $\varkappa$, the mean square rate of change in β. Hence it follows that

$$\tau = \frac{\bar{v}^3}{K\mu^2 D} = \frac{\bar{v}^3}{KG^2 m^2 D}. \qquad (168)$$

Some relaxation times estimated from Eq. (168) are listed in Table 4.

TABLE 4

Stellar system	$\bar{v}$, km/sec	$m/m_\odot$	D, pc^{-3}	τ, years
Star field near the Sun . .	30	1/2	0.12	$2 \cdot 10^{14}$
Globular cluster	10	1	5	$8 \cdot 10^{10}$
Open cluster	0.4	1	2	$3 \cdot 10^{7}$

The relaxation time of the star field around the Sun is thus exceedingly large, much greater than the conceivable age of the stars. Therefore during the entire lifetime of the Galaxy the irregular forces could not have had a substantial effect on the structure of the galactic region around the Sun and the local distribution of stellar velocities. It should be noted, however, that irregular forces are set up not by stars alone, but also by more bulky formations — such as star clouds and clusters, large clouds of diffuse matter, etc. If Eq. (168) is applied to star clouds, say, the factor mD in the denominator, which is equal to the density of matter, will remain as in the problem with individual stars, whereas the other factor, m, will increase by a factor of 10^7, giving a relaxation time of the order of 10^9. Proceeding from these considerations, Gurevich (1954), Spitzer and Schwarzschild (1953), and others are of the opinion that the galactic relaxation time is not particularly large compared to the star age. This problem has not been fully elucidated at this stage. Large-scale inhomogeneities

clearly reduce the relaxation time in the Galaxy, but this effect can hardly be considered with any accuracy, since in the derivation of Eq. (168) we regarded all the gravitating objects as point masses, whereas the stellar clouds and accumulations of diffuse matter are very extended objects.

The relaxation time in globular clusters is also large compared to the accepted ages of stars. It therefore seems that the globular clusters are not in a state of equilibrium in the irregular field.

In open clusters, on the other hand, the relaxation time is relatively short. This indicates that the stellar clusters have reached the state of equilibrium in the irregular field.

§164. Distribution function of the residual velocities and the rate of dissipation of stellar systems. Consider a nonrotating stellar system which is stationary in the regular field. Any such system should be spherical in shape, with zero centroid velocity at any point, so that the total velocities of the stars are equal to the residual velocities.

If the stellar systems were enclosed in gigantic sealed vessels preventing all dissipation of stars, it would follow from Boltzmann's theorem of statistical equilibrium in gases that stellar encounters should eventually lead to a Maxwellian velocity distribution.

Therefore in the stationary state the equation of balance

$$f(\beta)\int_{-\beta}^{\infty}\Phi(\beta,\ g)\,dg=\int_{0}^{\infty}f(x)\,\Phi(x,\ \beta-x)\,dx, \tag{169}$$

which states in mathematical terms that at any time the number of stars leaving the interval $[\beta,\ \beta+d\beta]$ is equal to the number of stars entering this interval, has a solution corresponding to a Maxwellian velocity distribution (163). The same result follows from the theory of drunkard's walk according to the limiting theorems of Markov processes.

Stellar systems, however, are not sealed in closed vessels, and stars with supercritical velocities escape from the system. The systems are therefore free from stars of supercritical velocities. This does not mean, however, that the distribution of subcritical velocities remains Maxwellian. For a quasistationary system this process of variation in β is described by a drunkard's walk model with an absorbing screen at the point β' corresponding to the square of the critical velocity.

In this model the distribution density near the absorbing screen is invariably less than the distribution density without absorption. The number of stars with near-critical velocities in stellar systems will therefore be less than the number prescribed by the Maxwellian distribution.

Since in a quasistationary stellar system the number of stars $N(\beta,\ t)\,d\beta$ in the interval $[\beta,\ \beta+d\beta]$ is a function of time, $N(\beta,\ t)$ being zero for $\beta>\beta'$, the equation of balance can be written in the form

$$N(\beta,\ t)\int_{-\beta}^{\infty}\Phi(\beta,\ g)\,dg=\int_{0}^{\beta'}N(x,\ t)\,\Phi(x,\ \beta-x)\,dx-\frac{dN(\beta,\ t)}{dt}. \tag{170}$$

Because of quasistationarity

$$N(\beta,\ t)=f(\beta)\,N(t),$$

where $N(t)$ is the total number of stars in the system. Eq. (170) therefore takes the form

$$f(\beta)\int_{-\beta}^{\infty}\Phi(\beta,\ g)\,dg=\int_{0}^{\beta'}f(x)\,\Phi(x,\ \beta-x)\,dx-f(\beta)\frac{1}{N(t)}\frac{dN(t)}{dt}\,. \qquad (171)$$

Eq. (171) is an approximate equation, since it is written for the system as a whole, whereas the value of β' is different at different points in the system. We are thus in fact dealing with some average solution of the system, i.e., the system is represented by some typical subsystem with average critical velocity. Then, according to the results of §164, $\beta'=4$.

The second term in the right-hand side of (171) is also expressed in terms of $f(\beta)$ and $\Phi(\beta, g)$, since the rate of dissipation is determined by

$$\frac{dN}{N}=dt\int_{0}^{4}f(\beta)\,d\beta\int_{4-\beta}^{\infty}\Phi(\beta,\ g)\,dg=\frac{\mu^2D\,dt}{\bar{v}^3}\int_{0}^{4}f(\beta)\,d\beta\int_{4-\beta}^{\infty}\Phi^*(\beta,\ g)\,dg. \qquad (172)$$

Eq. (171) is thus an integral equation for $f(\beta)$. It can be solved only by approximate methods. Therefore, seeing that the dissipation is a slow process, we can first ignore the second term in the right-hand side, find an approximate solution of the simplified expression, and then use this $f(\beta)$ to calculate (172) and in (171). For the values of β substantially less than β', $f(\beta)$ is not much different from (163). For values of β close to β', this difference is considerable. It is particularly large in problem B. The dispersion of the Maxwellian distribution for objects of zero mass is infinitely large, the function (163) becomes monotonically increasing and takes the form

$$f(\beta)=K\sqrt{\beta}, \qquad (173)$$

so that it can be normalized (in order to find K) only after introducing a cutoff, i.e., assuming that $f(\beta)$ has the form (173) for $\beta<\beta'$ and vanishes for $\beta>\beta'$. The introduction of an absorbing screen greatly lowers the values of $f(\beta)$ in the region near β'.

Table 5 lists the results of the approximate solution of (171) for problem A and compares these solutions with function (163). Since the solution depends on the values of g_0, two solutions are given: one for open, the other for globular clusters.

TABLE 5

β	lg $f(\beta)$		
	Maxwell distribution	open cluster	globular cluster
0	$-\infty$	$-\infty$	$-\infty$
0.3	$\bar{1}.89$	$\bar{1}.89$	$\bar{1}.89$
0.6	$\bar{1}.56$	$\bar{1}.56$	$\bar{1}.56$
0.8	$\bar{1}.10$	$\bar{1}.10$	$\bar{1}.10$
0.9	$\bar{2}.86$	$\bar{2}.85$	$\bar{2}.85$
0.95	$\bar{2}.75$	$\bar{2}.66$	$\bar{2}.66$
0.98	$\bar{2}.68$	$\bar{2}.49$	$\bar{2}.43$
0.99	$\bar{2}.65$	$\bar{2}.42$	$\bar{2}.27$
1.00	$\bar{2}.63$	$\bar{2}.34$	$\bar{3}.34$

Thus, if the distribution function over β is known, Eq. (172) gives the rate of dissipation

$$\frac{dN}{N} = \frac{\mu^2 D dt}{\bar{v}^3} H, \tag{174}$$

where

$$H = \int_0^4 f(\beta)\, d\beta \int_{4-\beta}^{\infty} \Phi^*(\beta, g)\, dg. \tag{175}$$

Application of Eq. (175) to case A gives the dissipation rate from a system of objects of average mass. For open clusters we get $H = 1.1$, for globular clusters $H = 4.0$. Inserting the functions $f(\beta)$ and $\Phi^*(\beta, g)$ for case B in (175) gives the rate of dissipation from a system of objects of vanishingly small mass. For open clusters we have $H = 43$, for globular clusters $H = 130$. The rate of dissipation from a system of objects of zero mass is thus one and a half orders of magnitude greater than the rate of dissipation for objects of average mass.

According to (174), an open cluster loses about 15% of its medium mass stars and virtually all the light stars during one complete revolution of the Galaxy, which is equal to $205 \cdot 10^6$ years for the Sun. During the same time a globular cluster loses only 0.008% of medium mass stars and 0.25% light stars. The dissipation of open clusters is thus very fast, whereas the dissipation of globular clusters is an extremely slow process. In galaxies, the effect of stellar encounters virtually rules out all possibility of dissipation. Dissipation may nevertheless occur following encounters with star clouds, star clusters, and other massive objects.

§165. The most probable distribution in phase space. As we have noted in the preceding (§154), a characteristic property of a system which is stationary in the regular field only is that the distribution of stars in phase space is not affected by the action of regular forces but does change under the action of irregular forces. Our problem now is to find how the system will change under the action of the irregular forces and to determine its final state. If there were no dissipation, this limiting state would clearly be the stationary state in both the regular and the irregular field. By analogy with statistical physics it is clear that it should correspond to the most probable distribution of stars in phase space. Following the procedure outlined by Ogorodnikov, we will determine the properties of this most probable distribution using the methods of statistical physics.

Dissipation effects are ignored. The velocities of stars are then limited by the critical velocities and the stellar system, which is regarded as bounded in normal space, also occupies a bounded volume in phase space. The finite phase volume is divided into s phase cells so that

$$1 \ll s \ll N,$$

i.e., on the one hand the number of cells is large and on the other hand each cell contains numerous stars.

Let γ_i be the volume of the i-th cell. Clearly,

$$\sum_{i=0}^{s} \gamma_i = \Gamma,$$

where Γ is the entire phase volume of the system.

Let now

$$\lambda_i = \frac{\gamma_i}{\Gamma}.$$

If the only restriction imposed on the system is the constancy of the number of stars (a natural assumption in the absence of dissipation)

$$\sum_{i=1}^{s} n_i = N = \text{const}, \tag{176}$$

where n_i is the number of stars in the i-th cell, we naturally assume that the probability of any star to be found in the i-th cell depends on the position of the cell in the phase volume and is equal to λ_i. According to the binomial distribution, the probability that for a random arrangement of stars in cells their numbers in the respective cells will be $n_1, n_2, \ldots, n_s$ is given by

$$P = \frac{N}{n_1!\, n_2! \ldots n_s!}\, \lambda_1^{n_1} \lambda_2^{n_2} \ldots \lambda_s^{n_s}. \tag{177}$$

Note that n_i and λ_i are related by the equality

$$n_i = \psi_i \lambda_i, \tag{178}$$

where ψ_i is the phase density in the i-th cell. If we were to calculate the most probable distribution under the sole restriction (176), we would find that n_i were proportional to λ_i, i.e., the stars would uniformly fill the phase volume giving constant phase density. In fact, however, two additional restrictions are imposed. The first is that in a stationary system the sum of the energies of the individual stars, defined by Eq. (98), is constant:

$$\sum_{i=1}^{s} n_i E_i' = \frac{m}{2} \sum_{i=1}^{s} n_i (\Pi_i^2 + \vartheta_i^2 + Z_i^2 + 2U_i') = \text{const}. \tag{179}$$

The second restriction is the constancy of the sum of the angular momenta of the stars about the rotation axis of the system:

$$m \sum_{i=1}^{s} n_i R_i \vartheta_i = \text{const}. \tag{180}$$

In writing Eqs. (179) and (180) we made use of the fact that all the stars in one phase cell can be assigned the same coordinates, so that each phase cell is represented by a single term in the sums. Since the phase density (the sought function) depends only on the integrals of motion of the stars, we need consider only restrictions imposed by integrals of motion of stars. Restrictions (176), (179), and (180) fall under this category and they in fact exhaust the field. Further note that Eq. (180) is exact, whereas Eq. (179) based on the virial theorem is fully applicable only to stationary systems or to systems with a linear variation of the moment of inertia. Since the system being considered is by assumption stationary in relation to regular forces, and the irregular forces produce slow variation of the system parameters, Eq. (179) is applicable to our discussion.

The problem now reduces to finding the maximum probability P, or equivalently the maximum $\ln P$, as a function of the variables $n_1, n_2, \ldots, n_s$ under constraints (177), (179), and (180). Consider the function

$$L = \ln P + k_1 \sum_{i=1}^{s} n_i + k_2 \sum_{i=1}^{s} n_i E_i' + 2k_3 \sum_{i=1}^{s} n_i R_i \vartheta_i, \tag{181}$$

where k_1, k_2, k_3 are Lagrange's indeterminate multipliers, which absorbed the stellar masses.

Equating to zero the partial derivatives of this function with respect to all the n_i, we find

$$\frac{\partial L}{\partial n_i} = -\frac{d \ln (n_i!)}{dn_i} + \ln \lambda_i + k_1 + k_2 E_i' + 2k_3 R_i \vartheta_i. \tag{182}$$

If the n_i are sufficiently large, we have

$$\frac{d \ln (n_i!)}{dn_i} = \frac{\ln [(n_i+1)!] - \ln (n_i!)}{(n_i+1)-n_i} = \ln (n_i + 1) \approx \ln n_i.$$

Inserting in (181), solving for n_i, and dropping the now superfluous subscript i, we get

$$n = \lambda e^{k_1 + k_2 E' + 2k_3 R\vartheta},$$

or, using (178),

$$\psi = e^{k_1 + k_2 E' + 2k_3 R\vartheta}. \tag{183}$$

To establish the physical meaning of this solution, we transform the power exponent in Eq. (183) as follows:

$$k_2(\Pi^2 + \vartheta^2 + Z^2 - 2U') + 2k_3 R\vartheta = k_2 [\Pi^2 + (\vartheta - \vartheta_0)^2 + Z^2 - 2U' - \vartheta_0^2], \tag{184}$$

where

$$\vartheta_0 = -\frac{k_3}{k_2} R. \tag{185}$$

Since the phase density is now seen to be a function of Π, $\vartheta - \vartheta_0$, and Z, we conclude that ϑ_0 is the centroid linear rotation velocity.

By Eq. (185) ϑ_0 is proportional to R. The centroid angular velocity is thus constant:

$$\omega = \frac{\vartheta_0}{R} = -\frac{k_3}{k_2} = \text{const}. \tag{186}$$

It also follows from (183) and (184) that the residual velocities follow a Maxwellian distribution which is the same everywhere throughout the system.

Thus, the most probable state of a stellar system is characterized by rigid-body rotation, Maxwellian velocity distribution, and constant kinetic temperature.

These results are quite acceptable. They show that the irregular forces in a stellar system are in a sense equivalent to viscosity, levelling out the differences in the angular velocities of adjoining layers in the system. Moreover, they act toward establishing a Maxwellian velocity distribution

at any point and also govern heat exchange processes between different parts of the system, eventually leading to uniform temperature. It should be remembered, however, that we ignored dissipation in our treatment. Our results therefore indicate the trend of events, whereas the system as a whole definitely does not rotate as a rigid body and is not isothermal. We moreover treated the system as bounded in space. This in its turn shows that the results are a priori inapplicable to the outermost regions of stellar systems.

The time for the system to attain the most probable state, i.e., to become stationary in relation to both the regular and the irregular field, we call the large relaxation time. Since the Galaxy and other spiral galaxies show differential rotation, they evidently have not reached the most probable state. The ordinary relaxation time — the time to establish the most probable distribution at a point — is apparently shorter than the large relaxation time.

§166. Dissipation of stellar systems. Dissipation of stellar systems due to the action of irregular forces was first considered by Ambartsumyan (1938). He determined the rate of dissipation of a nonrotating system, assuming that the fraction of stars dissipating during the relaxation time of the system is equal to the fraction of stars with supercritical velocities in the Maxwellian distribution. The critical velocity, according to §160, is equal to double the rms velocity of a star.

This argument is not entirely unobjectionable, since dissipation does not require establishment of a Maxwellian distribution in the entire velocity range: it only suffices that some stars acquire supercritical velocities. In this respect, the method proposed in §164 for the determination of dissipation rates is more reliable.

Comparison of the two methods has shown, however, that for medium mass stars Ambartsumyan's technique gives adequate results. The application of this method enabled Ambartsumyan to establish for the first time that open clusters decay fairly rapidly and that multiple Trapezium-type stars decay extremely rapidly.

In this section we will apply Ambartsumyan's technique to the more general case of rotating systems.

In 1950 Gurevich and Levin noted that owing to the great significance of close encounters, the velocities of stars mostly change in small portions. The velocity of a star escaping from the system thus in most cases exceeds the critical velocity by a very small amount, and the energy that these stars carry away is therefore virtually zero. This means that the total energy of a stellar system is virtually conserved in dissipation.

This important condition

$$E = \text{const} \tag{187}$$

provided a key to the study of the evolution of nonrotating gravitating systems, stellar systems included, in the course of their dissipation.

Eq. (187) is applicable to dissipating rotating stellar systems too, although in this case one must know how dissipation affects the total angular momentum of the system.

Note that escaping stars on the average carry away a larger angular momentum than the mean for the same number of stars in the system. Indeed, supercritical velocities are preferentially attained by stars whose

residual velocites make but a small angle with the centroid velocity. The angular momentum of these stars is clearly greater than the average angular momentum of stars from the same volume element. It therefore seems that the dissipating stars carry a certain excess angular momentum, in addition to the average angular momentum of the stars in the system. Let us now derive expressions for the average and the excess angular momentum carried away by dissipating stars.

Consider a small volume of a stationary system. Let R be its distance from the rotation axis of the system, M the angular momentum, ϑ_0 the linear velocity, n the number of stars in the volume. If ξ is the residual velocity of the star, and $\theta = \arccos \zeta$ is the angle between the residual velocity and the centroid velocity ϑ_0, the total velocity is given by

$$\omega^2 = \vartheta_0^2 + \xi^2 + 2\vartheta_0 \xi \zeta. \tag{188}$$

By averaging Eq. (188) and writing $\bar{\omega}$ and $\bar{\xi}$ for the rms total velocity and the rms residual velocity, we get

$$\bar{\omega}^2 = \vartheta_0^2 + \bar{\xi}^2. \tag{189}$$

The critical velocity of a given volume ω_k is well-defined, This velocity clearly is greater in the interior of the system, decreasing toward the periphery. Consider the nondimensional quantity

$$\alpha = \frac{\omega_k^2}{\bar{\omega}^2}. \tag{190}$$

If the velocity of a star is supercritical, it will escape from the system. It therefore follows from Eqs. (188)—(190) that for the stars escaping from the system

$$\vartheta_0^2 + \xi^2 + 2\vartheta_0 \xi x \geqslant \alpha (\vartheta_0^2 + \bar{\xi}^2). \tag{191}$$

We define a new factor

$$\gamma = \sqrt{\frac{3}{2}} \frac{\vartheta_0}{\bar{\xi}}. \tag{192}$$

γ can be used as a measure of the flattening of stellar systems. Clearly, the higher the ratio $\frac{\vartheta_0}{\bar{\xi}}$, the flatter is the system. In a nonrotating spherical system $\vartheta_0 = 0$, $\gamma = 0$, whereas in the limit of a perfectly flat system $\bar{\xi} = 0$, $\gamma = \infty$. Let further

$$y = \sqrt{\frac{3}{2}} \frac{\xi}{\bar{\xi}}. \tag{193}$$

In terms of these new functions, inequality (191) can be written as

$$y^2 + 2yx\gamma \geqslant (\alpha - 1)\gamma^2 + \frac{3}{2}\alpha. \tag{194}$$

Let the residual velocities of stars follow a Maxwellian distribution:

$$f(\xi, \theta)\, d\xi\, d\theta = \frac{3\sqrt{3}}{\sqrt{2\pi}} \frac{\xi^2}{\bar{\xi}^3} e^{-\frac{3}{2}\frac{\xi^2}{\bar{\xi}^2}} \sin\theta\, d\xi\, d\theta. \tag{195}$$

This corresponds to the following distribution of x and y:

$$f(y, x)\,dx\,dy = \frac{2}{\sqrt{\pi}}\,y^2 e^{-v^2}\,dy\,dx. \tag{196}$$

If the Maxwellian distribution were completely established, each escaping star would carry away angular momentum of

$$mr(\vartheta_0 + \xi x). \tag{197}$$

However, owing to the great importance of close encounters in the establishment of the actual velocity distribution, the velocities change in small portions and never greatly exceed the critical velocity. Therefore every supercritical velocity of the Maxwellian distribution should be replaced by a velocity pointing in the same direction whose magnitude, however, is equal to the critical residual velocity. This critical residual velocity ξ_k can be found by solving the limiting equality (194) for ξ:

$$\xi_k = \vartheta_0\left(\sqrt{x^2 + \alpha - 1 + \frac{3\alpha}{2\gamma^2}} - x\right). \tag{198}$$

Thus, the angular momentum carried away by each escaping body is

$$mr(\vartheta_0 + \xi k x). \tag{199}$$

To determine the fraction of stars dissipating from a given volume during the relaxation time, Eq. (196) should be integrated over the region defined by inequality (194) and

$$-1 \ll x \ll +1. \tag{200}$$

This gives

$$a(\gamma, \alpha) = \frac{2}{\sqrt{\pi}} \int_{-1}^{+1} dx \int_{\gamma\cdot\beta(x)}^{\infty} y^2 e^{-v^2}\,dy, \tag{201}$$

where

$$\beta(x) = \sqrt{x^2 + \alpha - 1 + \frac{3\alpha}{2\gamma^2}} - x. \tag{202}$$

To find the angular momentum dissipated during the relaxation time, we multiply (196) by (199) and n and integrate over the region defined by inequalities (194) and (200).

Using (198) and (200), we get

$$\frac{2}{\sqrt{\pi}} nmr\vartheta_0 \int_{-1}^{+1} [1 + x\cdot\beta(x)]\,dx \int_{\gamma\cdot\beta(x)}^{\infty} y^2 e^{-v^2}\,dy =$$

$$= nmr\vartheta_0 a(\gamma, \alpha) + nmr\vartheta_0 \frac{2}{\sqrt{\pi}} \int_{-1}^{+1} x\beta(x)\,dx \int_{\gamma\cdot\beta(x)}^{\infty} y^2 e^{-v^2}\,dy. \tag{203}$$

It is readily seen that if each dissipating star were to carry away angular momentum of $mr\vartheta_0$, which is equal to the average angular momentum per star in the given volume element, the entire dissipated momentum during the relaxation time would be equal to the first term in (203).

The second term in (203) thus represents the excess angular momentum dissipated by escaping stars. The origin of this excess angular momentum has been discussed in the preceding. Let $\varphi(\gamma, \alpha)$ denote the ratio of the excess momentum dissipated to the average momentum:

$$\varphi(\gamma, \alpha) = \frac{1}{a(\gamma, \alpha)} \cdot \frac{2}{\sqrt{\pi}} \int_{-1}^{+1} x\beta(x)\, dx \int_{\gamma \cdot \beta(x)}^{\infty} y^2 e^{-y^2} dy, \qquad (204)$$

where

$$\beta(x) = \sqrt{x^2 + \alpha - 1 - \frac{3\alpha}{2\gamma^2}} - x.$$

$\varphi(\gamma, \alpha)$ can be called the relative excess dissipated momentum.

Integrals (201) and (204) can be evaluated without quadratures: they are reduced by simple manipulations to expressions containing the probability integrals.

Table 6 tabulates the function $a(\gamma, \alpha)$, which represents the fraction of stars in a given volume escaping from the system during the relaxation time.

TABLE 6

α \ γ	0	0.8	1.2	1.6	2	3
2	$1.12 \cdot 10^{-1}$	$1.07 \cdot 10^{-1}$	$9.40 \cdot 10^{-2}$	$7.56 \cdot 10^{-2}$	$5.62 \cdot 10^{-2}$	$2.03 \cdot 10^{-2}$
3	$2.93 \cdot 10^{-2}$	$2.45 \cdot 10^{-2}$	$1.64 \cdot 10^{-2}$	$8.71 \cdot 10^{-3}$	$3.78 \cdot 10^{-3}$	$2.12 \cdot 10^{-4}$
4	$7.38 \cdot 10^{-3}$	$5.17 \cdot 10^{-3}$	$2.44 \cdot 10^{-3}$	$7.76 \cdot 10^{-4}$	$1.72 \cdot 10^{-4}$	$9.42 \cdot 10^{-7}$
5	$1.82 \cdot 10^{-3}$	$1.02 \cdot 10^{-3}$	$3.26 \cdot 10^{-4}$	$5.79 \cdot 10^{-5}$	$6.03 \cdot 10^{-6}$	$2.33 \cdot 10^{-9}$
6	$4.40 \cdot 10^{-4}$	$1.94 \cdot 10^{-4}$	$4.01 \cdot 10^{-5}$	$3.79 \cdot 10^{-6}$	$1.74 \cdot 10^{-7}$	$3.85 \cdot 10^{-12}$

The table shows that the dissipation during the relaxation time is greater for smaller α values, i.e., for smaller ω_k (assuming constant $\bar{\omega}^2$). This is a natural result, which reveals the predominance of the peripheral regions in dissipation. It should be remembered, however, that in the peripheral regions the relaxation time is greater owing to the small stellar density. Moreover, if the density is low, a given fraction $a(\gamma, \alpha)$ corresponds to a smaller number of stars. We should be careful not to overestimate the importance of the peripheral regions in dissipation processes.

The table also shows that $a(\gamma, \alpha)$ decreases with increasing γ, i.e., it is larger in nonrotating spherical systems and smaller in flat rotating systems. This seemingly unexpected result is associated with the fact that for small $\bar{\xi}$ the residual velocity ξ required for dissipation should exceed by a substantial factor the rms value ξ, and there are obviously very few such stars in a Maxwellian distribution.

It follows from Eq. (100) of §160 that the mean value of α for a stationary stellar system (in the regular field) is 4. Therefore, the proportion of stars dissipating during the relaxation time from a nonrotating system ($\gamma = 0$) is 0.0074 of the total number of stars. This estimate was first derived by Ambartsumyan. The dissipation of rotating systems, as we have noted, is slower.

Table 7 tabulates the relative excess momentum $\varphi(\gamma, \alpha)$ carried away in dissipation.

TABLE 7

α \ γ	0	0.8	1.2	1.6	2	3	∞
2	1.41	0.955	0.746	0.630	0.560	0.481	0.414
3	2.39	1.49	1.18	1.01	0.920	0.820	0.732
4	3.38	1.94	1.54	1.34	1.23	1.10	1.000
5	4.37	2.37	1.87	1.63	1.51	1.33	1.24
6	5.36	2.74	2.17	1.90	1.75	1.55	1.45

We see that the relative excess dissipated momentum is by no means small. Note that it is larger for spherical, slowly rotating systems than for flat fast spinning systems. It is also relatively large for stars dissipating from the interior of the system.

§167. **Evolution of dissipating stellar systems.** Proceeding from the conservation of the total energy of a dissipating system, Gurevich and Levin examined the evolution of nonrotating systems in the course of dissipation.

The constancy of total energy is written using (92) as

$$E = -\frac{Gm^2N^2}{4B\varrho} = const. \tag{205}$$

Since the shape of a nonrotating system does not change during dissipation, B remains constant. Of the various parameters entering Eq. (205), only N and ρ are variable in dissipation; therefore to preserve the equality we should have

$$\frac{\varrho}{\varrho_0} = \frac{N^2}{N_0^2}, \tag{206}$$

where nought subscribes the values of the parameters at some initial time.

Since the stellar density D is proportional to $\frac{N}{\varrho^3}$, it should vary as

$$\frac{D}{D_0} = \frac{N_0^5}{N^5}. \tag{207}$$

In dissipation D decreases. Eqs. (206) and (207) show that the radius of the system decreases faster than N, and the stellar density conversely increases very rapidly.

If a is the fraction of stars dissipating during the relaxation time τ, the change in the total number of stars in the system in time dt is

$$dN = -aN\frac{dt}{\tau}. \tag{208}$$

The relaxation time is given by Eq. (168):

$$\tau = \frac{\bar{v}^3}{KG^2m^2D},$$

and the rms velocity by Eq. (94):

$$\bar{w} = \sqrt{\frac{GmN}{2B\varrho}}.$$

In a nonrotating stationary system the centroid velocities ϑ_0 are zero and the total velocities of stars are equal to their respective residual velocities $\bar{w} = \bar{v}$. We thus have for the relaxation time

$$\tau = \frac{1}{(2B)^{3/2} K \sqrt{Gm}} \frac{N^{3/2}}{\varrho^{3/2} D}. \tag{209}$$

Inserting (209) in (208) and using (206) and (207), we get

$$dN = -\left[a \cdot (2B\varrho_0)^{3/2} K \sqrt{Gm} D_0 N_0^2\right] N^{5/2} dt. \tag{210}$$

Of the various parameters in brackets in (210), K is the only variable of dissipation. According to §163, however, the variation of K is not particularly pronounced. We may therefore ignore this variation, taking the entire product in brackets in (210) constant. The equation is readily integrated to give

$$N^{7/2} = N_0^{7/2} - \frac{7}{2} a (2B\varrho_0)^{3/2} K \sqrt{Gm}\, D_0 N_0^2 t, \tag{211}$$

and the integration constant is determined from the condition $N = N_0$ for $t = 0$.

If N is zero in Eq. (211), the corresponding t gives the total dissipation time t_d of the system. In the expression for t_d, we replace D_0 by the obvious expression

$$D_0 = \frac{3N_0}{4\pi\varrho_0^3},$$

and finally get

$$t_{d_0} = \frac{4\pi \sqrt{N_0 \varrho_0^3}}{21 B^{3/2} K a \sqrt{2Gm}}, \tag{212}$$

where the nought after t_d indicates that the total dissipation time is reckoned from the state with a total number of stars N_0 and a radius ρ_0. The dissipation time from any current state is obtained by substituting N and ρ for N_0 and ρ_0 in (212). Dividing the resulting equality by (212) and using (206), we get

$$\frac{t_d}{t_{d_0}} = \frac{N^{7/2}}{N_0^{7/2}}. \tag{213}$$

This equality shows that the dissipation of a nonrotating system is an accelerated process. It follows, in particular, that the dissipation of half the stellar population of the system takes a factor of $(8\sqrt{2} - 1)$ longer than the total dissipation of the remaining half.

Taking for an open cluster $N = 50$, $\rho = 2$ pc, $K = 50$, $m = m_\odot$ and for a globular cluster $N = 10^6$, $\rho = 20$ pc, $K = 130$, $m = m_\odot$, putting 0.0738 for the fraction of stars a dissipated during the relaxation time, and $B = 0.8$, we find from (212) that the total dissipation time of an open cluster is $3 \cdot 10^8$ years and for a globular cluster $8 \cdot 10^{11}$ years. An open

cluster thus could have completely dissipated during cosmogonically acceptable time.

The evolution of dissipating rotating systems presents a more formidable problem. Here the change in the angular momentum during dissipation should be taken into consideration. If to simplify the treatment we assume that the system on the whole evolves as any of its volume elements with average values of the relevant parameters (in particular, $\alpha = 4$), it can be shown that all the rotating stellar systems fall into two evolutionary sequences — spherical and disk. Systems with flattening $h = \frac{\varrho - \varrho_1}{\varrho}$ (ρ and ρ_1 are the semiaxes of the ellipsoid of revolution) less than 0.7 grow progressively more spherical during dissipation; their flattening steadily diminishes. These systems constitute the spherical evolutionary sequence. On the other hand, systems with flattening greater than 0.7 become progressively flatter in the course of their evolution, forming the flat (disk-type) evolutionary sequence.

The division into two evolutionary sequences is associated with the fact that, as we see from Table 7, the relative excess momentum carried away in dissipation is greater for slowly rotating, relatively spherical systems than for rapidly spinning ones. This markedly reduces the angular momentum of nearly spherical systems, improving their sphericity. The dissipated momentum of the highly flattened systems is relatively small and is on the whole insufficient to compensate for the acceleration of the system due to compression (see Eq. (206), which is valid for rotating systems too).

§168. General features in the evolution of stellar systems. If dissipation processes are ignored, the most probable state of a rotating gravitating system, according to §163, is rigid-body rotation with a Maxwellian distribution of residual velocities and constant dispersion at any point. Under the action of internal forces the system approaches this state, which is stationary in relation to both regular and irregular field.

Let us consider the approach of a nonstationary system to the stationary state in the regular field. On the whole, the regular forces alone cannot effect this transformation. Indeed, if an additional free body is placed at some distance from a spherical system which is stationary in the regular field, it will oscillate indefinitely in the regular field approximation, passing back and forth through the spherical system. The system including this body will therefore remain nonstationary. If, however, irregular forces are taken into consideration, the oscillating body will experience repeated close encounters with other stars in the system and in a time comparable with the relaxation time it will be absorbed and become an ordinary member of the spherical system, an indistinguishable part of the established velocity distribution of the entire system.

The combined effect of regular and irregular forces thus brings a system, originally stationary in the regular field, to a stationary state in the irregular field as well.

In particular cases, however, which are of considerable practical importance, the system can be brought to a stationary state in the regular field by the action or regular forces only. These forces produce "mixing" of the system, and the time of "mixing" is large compared to the mean time for a star to cross the system.

When "mixing" is terminated, the transition to a stationary state in regular and irregular field is effected by the action of irregular forces only.

On the whole, the time for the establishment of the stationary state (relaxation time) is different for different parts of the system, so that the entire system does not become stationary at one time. Eventually, however, after a period of time greater than the largest of the different relaxation times, the system will become stationary. We call this state stationary at any point. In this state the angular rotation velocity of the centroid and the dispersion of the residual velocities at various points of the system are in general not constant.

This is not the most probable state. It is stationary in the regular field, but only quasistationary in the irregular field. The imperfect stationarity in the irregular field is associated with viscosity and "thermal diffusion": these factors produced by the irregular forces gradually level off the differences in the angular centroid velocities and the dispersion of residual velocities throughout the system. This process* is terminated when the most probable state is established, or in other words when the system has become stationary on the whole.

The state of the system is essentially influenced by dissipation. The steady action of irregular forces will impart supercritical velocities to a certain proportion of objects in the system, which will consequently escape. The process of dissipation is continuous, since after the supercritical objects have escaped from the system, other objects acquire supercritical velocities and escape from the system, and so on. This process causes slow but steady departure from stationarity. On account of dissipation, the parameters characterizing the state of the system are clearly different from those obtaining without dissipation. In particular, because of dissipation, the most probable state of the system is not perfectly isothermal nor is its rotation a perfect rigid-body rotation.

Ambartsumyan used the term quasistationary state to describe systems which have reached equilibrium but continue changing due to dissipation. We will correspondingly use the terms locally quasistationary state and global quasistationary state.

In a global quasistationary state, dissipation is the only factor responsible for departure from perfect stationarity. Although dissipation produces differences in the angular centroid velocities and the dispersion of residual velocities, these effects are clearly a result of dissipation.

In a locally quasistationary state the departure from perfect stationarity is due not only to dissipation alone, but also to the independently existing differences between angular centroid velocities and dispersions of residual velocities at different points.

A characteristic property of the global quasistationary state is that knowledge of the principal general parameters of the system — mass (equal for all the objects), number of objects, energy and angular momentum of the system — uniquely determines the state of the system. Ogorodnikov called these systems dynamically determinate. Note that only globally quasistationary systems are dynamically determinate.

Let us sketch a general outline of the evolution of rotating self-gravitating systems.

* L. E. Gurevich (1954) interprets this process as transfer of angular momentum from the interior to the periphery.

The initial state of the system is determined by the conditions of its formation. If the initial state was stationary in the regular field and nonstationary in the irregular field, each volume element will reach quasistationary state after the elapsing of its characteristic relaxation time, and after a period greater than the largest of the relaxation times the system will become quasistationary at every point (locally quasistationary state).

In stellar systems the relaxation time generally increases with distance from the rotation axis; the locally quasistationary state is therefore first established in the central region, and then the sphere of local quasistationarity gradually expands toward the periphery. If at the time of formation the system was nonstationary in the regular field as well, it will first rapidly approach its stationary state in the regular field (due to "mixing" by regular forces), and then progress to local stationarity by combined interactions of regular and irregular forces.

A locally quasistationary system then evolves to global quasistationary state, with conditions close to isothermal and rigid-body rotation. The duration of this stage, which in §163 we called the large relaxation time, is largely dependent on the spin of the system. The large relaxation time increases with increasing population and increasing γ — the ratio of the mean linear centroid velocity to the mean residual velocity of the objects in the system. If the system does not rotate ($\gamma = 0$), locally quasistationary state and global quasistationary state are attained simultaneously, i.e., the large relaxation time is zero. If the system spins fast ($\gamma \gg 1$), the adjacent annular regions are largely autonomous, the levelling out of centroid velocities and velocity dispersions is a slow process, and the large relaxation time may exceed the relaxation time at a point many orders of magnitude.

Global quasistationary state is also first established at the center and then propagates to the periphery. A substantial region at the center may reach a global quasistationary state even before the entire system has become locally quasistationary. The large relaxation time for this region (which is always simply-connected and symmetric relative to the symmetry axis and the plane of symmetry of the entire system) depends on the value of γ at its boundary.

Gurevich (1954) has shown that the globally quasistationary region forming at the center of the system produces a definite central bulging.

Let us establish the characteristic properties of the residual velocity distribution function at various stages of system evolution.

In nonstationary systems, no restrictions are imposed on the residual velocity distribution function.

If we consider those parts which are stationary in the regular field, the distribution of the small residual velocities is such (see §159) that the projections of the velocities on the plane of symmetry of the system follow an elliptic distribution.

If the region being considered is at the same time nonstationary in the irregular field, the total velocity (including the component along the symmetry axis of the system) does not show an ellipsoidal distribution and the dispersion of the residual velocities in this third direction is in general arbitrary (though, of course, bounded). The ratio of the residual velocity dispersions in the direction to the center of the system, along the centroid

velocity vector, and along the symmetry axis is therefore

$$1 : \frac{\sigma_2^2}{\sigma_1^2} : \frac{\sigma_3^2}{\sigma_1^2} = 1 : \frac{-B}{A-B} : \frac{\sigma_3^2}{\sigma_1^2}. \tag{214}$$

If the region being considered is stationary (quasistationary) in both regular and irregular fields, the distribution is determined by the larger of the two quantities — the relaxation times τ at the points of the given region or the period

$$T = \frac{\pi}{\sqrt{B(B-A)}}$$

for the tip of the projection of the residual velocity vector to describe an ellipse in the regular field (see §159).

If τ is large compared to T, the residual velocities follow an ellipsoidal distribution and according to §158 the ratio of the dispersions of velocities in the three directions (in the present case, the ratio of the ellipsoid semi-axes squared) is equal to

$$1 : \frac{-B}{A-B} : 1. \tag{215}$$

If now τ is small compared to T, i.e., the effect of the irregular forces becomes noticeable sooner than the effect of the regular faces (an uncommon case in stellar systems), the velocity distribution is Maxwellian spherical, i.e., the ellipsoid axis ratio is

$$1 : 1 : 1. \tag{216}$$

If the region is quasistationary on the whole, we have $A = 0$ and the ellipsoid axis ratio is equal to (216) irrespective of which of the two times — τ or T — is larger.

Observations show that the dispersion ratio of the residual velocities in the three directions for the nearby stars satisfies condition (214). The ratio $\frac{-B}{A-B}$ derived from galactic rotation data (see §150) is found to be 0.44, and $\frac{\sigma_2^2}{\sigma_1^2} = 0.43$. Also $\frac{\sigma_3^2}{\sigma_1^2} = 0.25$. It therefore seems that the galactic region around the Sun is stationary in the regular field and nonstationary in the irregular field.

Another observational fact which is indicatory of the evolutionary stage of stellar systems is the presence of bulging nuclei in spiral galaxies similar to our Galaxy; these nuclei invariably rotate as a rigid body. The spectral lines from the central regions of galaxies are inclined but straight, which is a definite proof of rigid-body rotation. The convexity of the nuclei shows that they evolve as systems of the spherical evolutionary sequence and have by now covered a substantial part of their evolutionary path.

We thus conclude that our Galaxy at present has evolved to a stage when it is stationary in the regular field, locally quasistationary state has been established throughout a considerable part of the system but has not reached the periphery (nor the neighborhood of the Sun), and the central part is quasistationary as a whole.

REFERENCES TO CHAPTER XX

1. AGEKYAN, T.A. The evolution of rotating systems of gravitating objects.— Astron. Zh., 35:26. 1958.
2. AGEKYAN, T.A. Probability of stellar encounter with a given change in absolute velocity.— Astron. Zh., 36:41. 1959.
3. AGEKYAN, T.A. Velocity distribution function and rate of dissociation in gravitating systems.— Astron. Zh., 36:183. 1959.
4. AMBARTSUMYAN, V.A. On the dynamics of open clusters.— Uchenye Zapiski, Leningrad Univ., No.2, Issue 4. 1938.
5. GUREVICH, L.E. Evolution of stellar systems.— Voprosy Kosmogonii, 2:151. 1954.
6. GUREVICH, L.E. and B.Yu. LEVIN. Evolution of gravitating systems.— Doklady AN SSSR, 70, No.5:781.
7. KUZMIN, G.G. A third integral of motion of stars and the dynamics of stationary Galaxy.— Publ. Tartu Astron. Observ., 32, No.5:332. 1953.
8. KUZMIN, G.G. On the value of the dynamic parameter C and the density of matter around the Sun.— Publ. Tartu Astron. Observ., 33, No.1:3. 1955.
9. OGORODNIKOV, K.F. Elements of dynamics of rotating stellar systems.— Uspekhi Astron. Nauk, 4:3. 1948.
10. OGORODNIKOV, K.F. Dinamika zvezdnykh sistem (Dynamics of Stellar Systems).— Fizmatgiz. 1958.
11. PARENAGO, P.P. The gravitational potential of the Galaxy, part I.— Astron. Zh., 27, No.6. 1950; part II — Astron. Zh., 29, No.3. 1952.
12. PARENAGO, P.P. Kurs zvezdnoi astronomii (A Course in Stellar Astronomy), 3rd ed.— Gostekhizdat. 1954.
13. SAFRONOV, V.S. The density of matter in the Galaxy and near the Sun.— Astron. Zh., 29, No.2. 1952.
14. CHANDRASEKHAR, S. Principles of Stellar Dynamics.— University of Chicago. 1942.
15. JEANS, J.H. Problems of Cosmogony and Stellar Dynamics.— Cambridge, 1919.
16. LINDBLAD, B. Die Milchstrasse.— Handbuch der Astrophysik, Bd. 5, 1 — 2. 1933.
17. LINDBLAD, B. Arkiv for Matem., Astronomi och Fysik, 20A, No.7. 1927.
18. OORT, J.H. Dynamics of the galactic system in the vicinity of the Sun.— Bull. Astron. Inst. Netherl., 4, No.159. 1928.
19. OORT, J.H. The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems.— Bull. Astron. Inst. Netherl., 6, No.38. 1932.
20. PAHLEN, E. Lehrbuch der Stellarstatistik.— Leipzig, 1937.
21. SPITZER, L. and M. SCHWARZSCHILD. The possible influence of interstellar clouds on stellar velocities, I.— ApJ 114, 385—397. 1951; II.— ApJ 118, 106—112. 1953.

T.A. AGEKYAN

Chapter XXI

STAR CLUSTERS

§ 169. Introduction. A considerable proportion of stars in the Galaxy are grouped into discrete formations of various sizes. This feature of our stellar system is also characteristic of other spiral and irregular galaxies and it is probably observed in elliptic galaxies as well.

In Chapters III through V we described double and multiple stars, which are in fact physical systems comprising two or several stars. Another type of star combinations are the star clusters—systems containing more numerous stars and characterized by a more complex structure. Star clusters may include double and multiple stars as constituent elements.

We distinguish between two main types of stars clusters: o p e n and g l o b u l a r . These two types greatly differ in their physical properties.

The clusters are generally designated according to the New General Catalogue of Nebulae and Clusters, compiled by Dreyer in 1888. For example, an object listed under No. 6649 in Dreyer's catalogue (a cluster in this case) is designated NGC 6649. In 1895 and 1908 Dreyer published supplements to his main catalogue, the so-called Index Catalogues; the corresponding objects are designated IC. Messier's catalogue published in 1784 is also used (the corresponding symbol being M), but this catalogue is being gradually abandoned. Melotte's catalogue (Me) published in 1915 is used even more seldom.

The most famous clusters have special proper names, such as Pleiades, Hyades, h and χ Persei.

§ 170. Open star clusters. Open star clusters are sometimes called galactic clusters. The latter term emphasizing the pronounced galactic concentration of the open clusters is most unfortunate, since globular clusters are also galactic objects. The term "open cluster" was introduced in the 1780's by W. Herschel, who first studied these objects. Figure 182 is a photograph of the open cluster NGC 2362, taken through the 100-in. telescope.

The number of observed stars in open clusters ranges from a few tens to a few hundreds. NGC 2099 contains about 1000 stars. The number of stars actually contained in a cluster is substantially greater than the number of observed stars, especially in distant clusters, since only stars of relatively high luminosity are visible.

Open clusters are observed as patches of increased surface density in the apparent distribution of stars. The boundaries of these patches are highly uncertain. Naked-eye estimates of the apparent size of open clusters

invariably give too low results. The method of star counts is more reliable. Applied to stars up to $9^{m}.0$ in the Pleiades, it gives the results listed in Table 1.

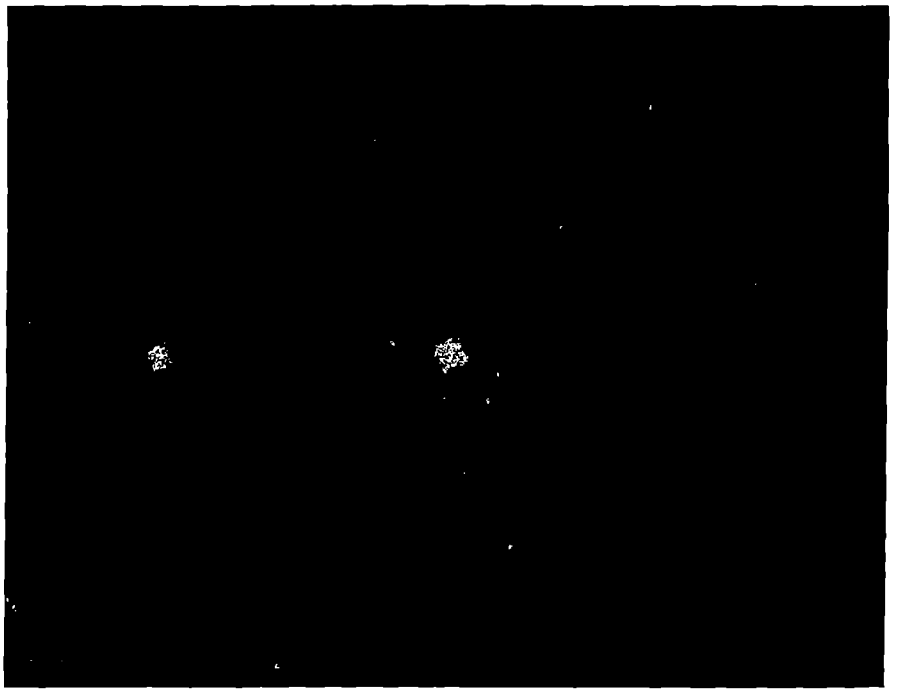


FIGURE 182

The second column in Table 1 gives the count of stars in annular zones centered at the point of maximum condensation of the surface star density; the radii of the successive rings are given in the first column. The third column shows that the surface star density at first decreases rapidly with distance from the center, but then it levels off. The mean star density for the last two rings is 3.8 stars per sq. deg and is entirely attributable to the field stars. We can thus calculate the expected number of field stars and cluster stars in each ring.

TABLE 1

Distance from the center	Number of stars	Number of stars per 1 sq. deg.	Probable number	
			field stars	cluster stars
0°—1°	63	20.0	12	51
1°—2°	56	5.9	36	20
2°—3°	79	5.0	60	19
3°—4°	87	4.0	84	3
4°—5°	104	3.7	108	—4

The cluster thus extends approximately over 3° from the center and the fifth column shows that the probable number of stars up to 9^m in the cluster is 90.

The star counting technique is not concerned with the individual members of the cluster. To identify the cluster stars, we primarily require the proper motions of the stars. The members of the cluster should have similar proper motions. We also study the radial velocities, since they should be almost identical for the stars of a cluster. Finally, the color index — apparent magnitude diagram is used. This diagram should have precisely the same form as the color index — absolute magnitude diagram, but it is displaced along the ordinate depending on the actual distance of the cluster. It is also slightly shifted along the horizontal axis because of selective absorption. The stars of the cluster fall along the main sequence in the diagram; if the cluster contains some red giants, the diagram will have a giant branch also. Field stars noticeably deviate from these two branches.

A combination of the three selection techniques enables us to identify with fair certainty the individual members of the cluster: the application of the color index — apparent magnitude diagram selects those stars which lie inside the cluster and consideration of proper motions and radial velocities guarantees common spatial velocity of the stars and the cluster.

The assumption that all the stars of a given cluster have equal spatial velocities is naturally not exact. The stars in the cluster also have internal motions. However, studies of open clusters have shown that the velocities of these internal motions are small compared to the spatial velocity of the center of inertial of the cluster relative to the Sun.

The composition of the open clusters is represented by Population I stars (Baade). The main sequence is always clearly delineated in the color index — apparent magnitude diagram. The giant branch is generally absent. When it is observed, it contains only few representatives.

Open clusters often contain stars with metallic lines. There are numerous Be stars with emission lines. Some indications of white dwarfs are available. Double stars are present. As regards variable stars, until recently it was generally held that open clusters contained no variables. Lately, however, some open clusters were found to contain about a dozen cepheids. T Tauri variables, stars with variable spectra, and eclipsing variables are also observed. RR Lyrae stars are conspicuous in their total absence. Note that variables in general constitute a minute fraction of all the stars. Since open clusters are relatively low-population objects, they cannot be expected to contain variable stars in great numbers.

The color index — apparent magnitude diagram for star clusters is much more certain than the color index — absolute magnitude diagram for galactic field stars. This is so because distance errors for the galactic field stars lead to an artificial increase in the dispersion of the absolute magnitudes of stars of a given color index. For clusters this source of dispersion vanishes, since all the stars are virtually at the same distance from the observer. The significance of the color index — apparent magnitude diagrams of star clusters therefore extends far beyond the limits of the particular problem of star clusters. They can be applied to study in greater detail the properties of the color index — absolute magnitude diagram and are thus of cosmogonic importance.

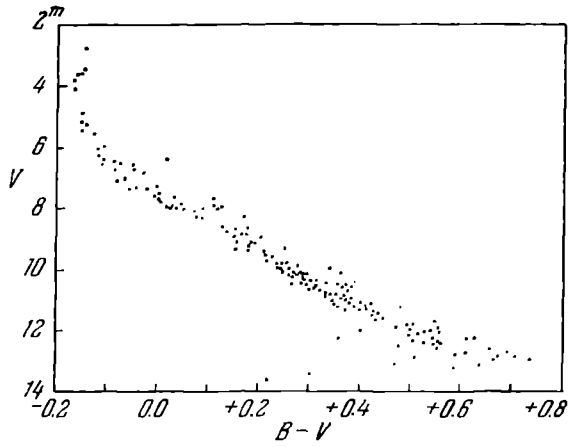


FIGURE 183

Figure 183 is the color index — apparent magnitude diagram of the Pleiades in the Johnson-Morgan system, constructed by Mitchell and Johnson in 1957. Note that the diagram does not show a single red giant. The main sequence stars dominate, and they are spread fairly uniformly over the entire observed section. No red dwarfs are observed, apparently because of their faintness.

The diagram for NGC 6530 is somewhat different (Walker, 1957; Figure 184). Here there is a trace of the red giant branch. At the same time, the stars of the earliest type have zero color index, i.e., these are A0 stars. NGC 6530 thus contains no hot giants.

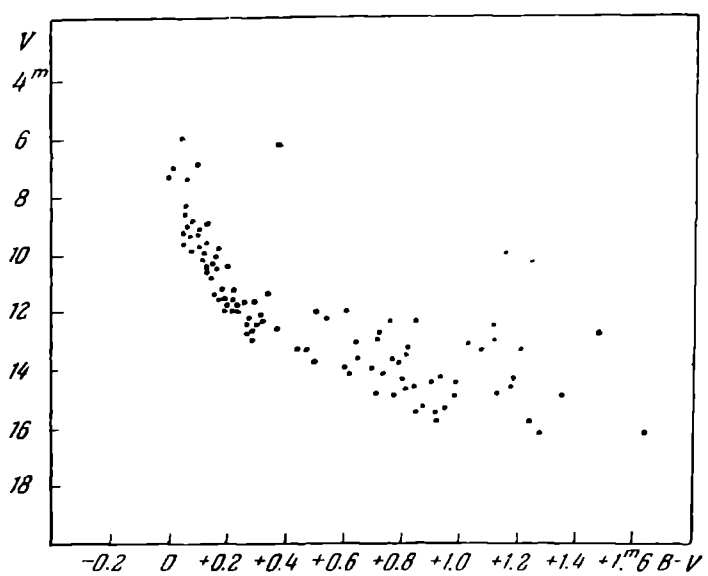


FIGURE 184

Figure 185 is the spectrum — absolute magnitude diagram of Pleiades constructed by Mendoza (1956).

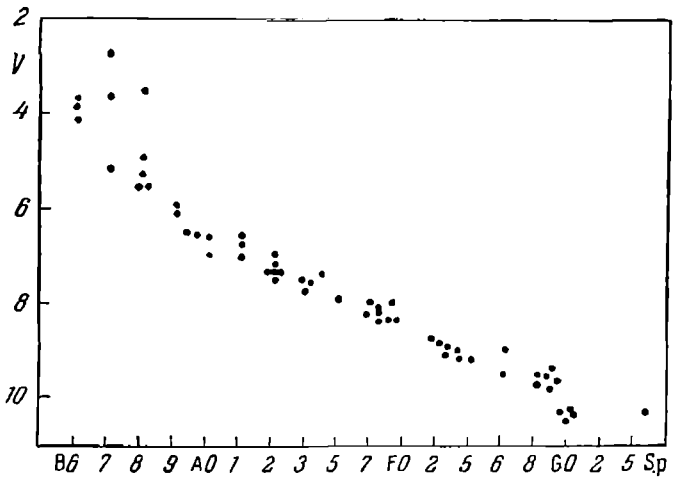


FIGURE 185

W. Herschel was the first to classify the open cluster. His classification was based on structural features and morphological differences. It covered nebulae, as well as open clusters proper.

A more detailed classification of open clusters was proposed by Trumpler. It is based on stellar composition and structure of the open clusters. The stellar composition is specified by Arabic numerals and lower-case Latin letters. The numeral 1 signifies that the giant branch is absent altogether, 2 corresponds to clusters with an inconspicuous giant branch, 3 to clusters with a majority of red and yellow stars. The Latin letter identifies the spectral type where the main sequence begins. For example, the Pleiades are classified as 1b, since they have no red giants and the main sequence starts with spectral type B.

The distribution of 100 open clusters in Trumpler's classification is given in Table 2.

TABLE 2

	o	b	b — a	a	a — f	f	Total
1	7	24	5	3	0	0	39
1—2	3	15	10	3	0	0	31
2	0	1	5	18	1	1	26
2—3	0	0	0	3	0	0	3
3	0	0	0	1	0	0	1
Total	10	40	20	28	1	1	100

The concentration of large numbers along the diagonal of this table shows that clusters with O—B stars contain virtually no red giants. These clusters are the most numerous. If the main sequence starts with spectral type A, the cluster contains some red giants.

In open clusters, hot giants and cold giants as if avoid one another. This conclusion is of great cosmogonic significance. It implies that either the conditions prevailing in each cluster are conducive to the formation of one type of giants only or clusters form originally with giant population of one certain type, which eventually evolves into giants of the other type. The latter point of view is currently favored by numerous American astronomers, who maintain that the evolution of hot giants into red giants can be traced from detailed observations of the open clusters.

The hot giants in open clusters are of higher luminosity than the red giants. There are no red supergiants in open clusters, whereas white and blue supergiants are fairly frequent. Note, however, that the great number of o and b clusters is partly attributed to observational selection. Clusters of this type are the easiest to detect.

The structure characteristics of open clusters in Trumpler's classification are designated by Roman and Arabic numerals and Latin letters:

I — a cluster with pronounced central concentration;
II — a cluster with moderate central concentration;
III — a cluster without central concentration;
IV — a cluster almost merging into the background;
1 — all visible stars have equal brightness;
2 — a gradual transition from bright to weak stars is observed;
3 — the cluster comprises a group of bright stars and a group of weak stars;
p — a poor cluster (less than 50 stars);
m — a moderate cluster (from 50 to 100 stars);
r — a rich cluster (over 100 stars).

The Pleiades, in Trumpler's classification, are designated b1IIr.

Trumpler's classification of open clusters has one fundamental shortcoming: the apparent structure characteristics are more likely to combine into one group clusters occurring at equal distance from the observer and having similar visibility conditions, rather than physically similar clusters. Thus, distant clusters show a more pronounced apparent concentration of stars toward the center than the nearby clusters do. In distant clusters only the brightest stars are observed, and the cluster will therefore appear as poor with visible stars of approximately uniform brightness.

A new classification of open clusters was developed by Markaryan. Retaining the basic idea of classification according to stellar composition, he considered three types of clusters, O, B, and A. O clusters are those with the brightest stars of spectral types O and B0; B clusters are those with the brightest stars of types B1—B5; A clusters are those with the brightest stars of types B6—A3.

Rejecting the visible structure features which may be the accidental result of distance and visibility condtions, Markaryan used in the classification of O clusters the fine structure features studied at the Byurakan Observatory. According to his system, O_R clusters are those which contain as a nucleus one or several multiple stars of the Trapezium in Orion type; O_n include clusters where straight or arched chains of stars are observed; O_p are clusters divided into two parts by a star-free corridor; clusters associated with luminous gaseous nebulae are designated as O_N.

If a cluster combines several of the above features, the symbol O is followed by a string of the corresponding subscripts.

Four methods are available for the determination of distances of open clusters: the method of stellar luminosities, the method of diameters, the method of radial velocities, and the method of spectroscopic parallaxes. The first two methods are the most commonly used.

To apply the luminosity method, we have to determine the spectral type and the apparent magnitude of as many stars in the cluster as possible. Having separated between the dwarfs and the giants and eliminated the star fields to the best of our ability, we proceed to calculate the mean apparent magnitude $\overline{m}_s$ of stars of each spectral type. This magnitude is compared with the average absolute magnitude $\overline{M}_s$ of stars of the corresponding spectral type. The distance is calculated from the relation

$$\lg r = \frac{1}{5}[\overline{m}_s - \overline{M}_s - A(r)] + 1. \qquad (1)$$

The absorption $A(r)$ is best determined from the color excesses of the same stars. Parenago's formula can also be used:

$$A(r) = \frac{a_0\beta}{\sin b}\left(1 - e^{-\frac{r\sin b}{\beta}}\right).$$

The method of diameters is applied to distant open clusters. It is based on the assumption that clusters of the same type have the same linear dimensions. The application of this method naturally requires knowledge of the average diameter D_c of clusters of a given type. D_c can be determined for the near clusters, where the stellar spectra are determined with higher reliability and where the absorption error is not very significant. If D_c is known, the distance to a cluster of the same type can be calculated from the relation

$$r = 3438\frac{D_c}{d}, \qquad (2)$$

where D_c is in parsecs and the angular diameter d is in minutes of arc.

The method of diameters was widely used by Trumpler in whose opinion the main advantage of the method was that, once D_c has been found, the distance to the clusters is no longer distorted by interstellar absorption. Barkhatova has shown, however, that this assumption is not entirely correct. Because of the patchy absorption of light, individual parts of clusters may suffer greater attenuation, and the angular size of the cluster will come out too low. This effect is clearly more pronounced at greater distances.

The distances of open clusters can be determined by a combination method. For each cluster we have

$$5\lg r_i + ar_i = \overline{m}_s - \overline{M}_s + 5,$$
$$\lg D_c - \lg r_i = \lg d - 3.536.$$

Examining n clusters, we obtain $2n$ equations. The right-hand sides are obtained from observations. The left-hand sides contain n unknowns r_i and k unknowns D_c (according to the number of clusters) and an unknown absorption coefficient a. Since $2n > n+k+1$, we have a redundant system

which can be solved by the least squares method. Applying this method, Trumpler first determined the mean absorption coefficient in directions close to the galactic plane.

The determination of the distances of open clusters from radial velocities uses galactic rotation data. As a first stage, the equation

$$v_r = k_1 + k_2 r \cos^2 b + Ar \cos^2 b \sin 2(l - l_0) \qquad (3)$$

(see §149) is applied to clusters (more than four) whose radial velocities are known and whose distances have been determined by some alternative method. Then the least squares method is applied to determine the rotation parameters k_1, k_2, A, and l_0 of the system of open clusters. Once the parameters have been determined, Eq. (3) is used again to determine the distance of any open cluster with known radial velocity.

Since the coefficient k_2 (the k effect) is small compared to Oort's constant A, the distances determined using Eq. (3) for clusters at galactic longitudes close to 0, 90, 180, and 270° (reckoned from the direction to the center of the Galaxy) are unreliable.

The diameters of open clusters range between 1.5 and 15 pc. Trumpler derived the frequency of diameters for 100 clusters, which is listed in Table 3.

The table shows that the great majority of open clusters have diameters between 2 and 7 pc. The modal diameter is ≈4 pc, the mean diameter ≈5 pc. Clusters with high concentration are on the average smaller. The average diameters of types I, II, III, and IV according to Trumpler are 4.0, 4.4, 5.9, and 10.3 pc, respectively.

TABLE 3

Diameters	Number of clusters	Diameters	Number of clusters
0—1	0	7—8	4
1—2	1	8—9	1
2—3	18	9—10	4
3—4	25	10—11	1
4—5	21	11—12	3
5—6	13	12—13	1
6—7	7	13—14	0
		14—15	1

The average spatial star density in open clusters is from 1 to 5 stars per 1 pc^3, i.e., 1—1.5 orders of magnitude greater than the galactic field density near the Sun. The density in the central parts of open clusters is even higher, reaching for NGC 6705, say, up to 80 stars per 1 pc^3. It should be further remembered that not all the stars actually contained in the cluster are visible, so that the true star density is higher than the observed density. Since open clusters are made up of stars of comparatively high luminosities, the mean mass of stars in these clusters is approximately equal to the Sun's mass (around the Sun, the mean stellar mass is approximately 0.42 of the solar mass). The mean density of matter in open clusters is therefore 1—5 solar masses per 1 pc^3.

The number of known open clusters is currently over 500. They form a distinct disk subsystem. The degree of galactic concentration in the open clusters can be characterized by their distribution according to galactic latitude, based on Hogg's data (1959, see Table 4).

The prevalence of open clusters with negative galactic latitudes is apparently associated with the fact that the Sun lies above the plane of the Galaxy.

TABLE 4

Galactic latitude	Number of clusters	Galactic latitude	Number of clusters	Galactic latitude	Number of clusters
$-90° - -30°$	4	$0° - 40°$	21	<0.5	46
$-30° - -15°$	13	$40° - 80°$	52	$0.5 - 1.0$	131
$-15° - -6°$	43	$80° - 120°$	74	$1.0 - 1.5$	115
$-6° - 0°$	231	$120° - 160°$	40	$1.5 - 2.0$	74
$0° - +6°$	167	$160° - 200°$	72	$2.0 - 2.5$	33
$+6° - +15°$	42	$200° - 240°$	71	$2.5 - 3.0$	34
$+15° - +30°$	10	$240° - 280°$	56	$3.0 - 4.0$	19
$+30° - +90°$	4	$280° - 320°$	62	>4.0	33
		$320° - 360°$	65		
Total	516	Total	516	Total	439

Despite the asymmetric position of the Sun in the Galaxy, the open clusters show a fairly uniform distribution in longitude. This shows that the open clusters have a negligible concentration toward the galactic center (as is proper for objects of disk subsystems) and that they are fairly numerous toward to periphery of the Galaxy.

Let us consider the distribution of open clusters over distances. Since the disk subsystem is extremely flattened, the volumes corresponding to the distance intervals selected in the table may be treated as annuli in a ratio of 1:3:5:7:9:11:28:∞. We see that the number of clusters in the first two annuli is almost proportional to the corresponding volumes, whereas in the next intervals the number of clusters shows an even increasing deficiency. This effect can be explained by the fairly uniform distribution of clusters near the plane of the Galaxy, assuming that all the clusters up to distances of 1 kpc have already been discovered. Among the more distant clusters, the proportion of undiscovered objects increases with increasing distance. Since within a radius of 1 kpc there are 187 clusters and the radius of the Galaxy is approximately 12 kpc, the total number of open clusters can be estimated as $187 \times 12^2 \approx 27{,}000$.

TABLE 5

NGC	l	b	R'	r_{pc}	m	m^*	N
457	94.4	−3.9	10	1400	8.0	8.6	100
1039	111.7	−14.8	18	400	5.8	8.6	80
Praesepe	114.8	−6.2	240	150	2.3	5.0	40
Pleiades	134.7	−22.2	100	120	1.6	4.2	130
2099	145.3	+4.5	20	800	6.2	9.7	150
2168	154.3	+3.6	40	700	5.6	9.0	120
2632	173.3	+34.0	95	200	3.9	7.5	75
3228	248.6	+4.8	20	500	6.4	10.1	?
3114	250.9	−3.6	30	500	4.5	7.0	100
3532	257.4	+1.5	60	400	3.4	8.1	130
6405	324.3	−2.0	25	400	4.6	8.3	50
6705	355.0	−4.3	10	1700	6.8	12.0	200

The integrated absolute magnitude of the open clusters is hardly dependent on the total number of stars. It is actually determined by the luminosity of the brightest stars in each cluster. The frequency curve of absolute magnitudes of open clusters is peaked near $-3^m.5$; the curve deviates from the normal distribution toward an excess of high-luminosity clusters.

Table 5 lists the basic data for 12 open clusters.

In Table 5 R' is the radius of the cluster in minutes of arc, r_{pc} is the distance in parsecs, m is the integrated stellar magnitude of the cluster, m^* is the apparent stellar magnitude of the brightest star, N estimated number of the visible stars in the cluster.

§ 171. **Moving clusters.** Let the spatial velocity vector of the center of inertia of an open cluster relative to the Sun be V. The point of the sky to which the vector V points is called the radiant of the cluster. The proper motions of all the stars in the cluster should be directed to the radiant if the cluster moves away from the Sun. If the cluster approaches the Sun, the proper motion vectors diverge from the antiradiant. Let A and D be the equatorial coordinates of the radiant. The componets of V in rectangular coordinates in the equatiorial system are then given by

$$X = V\cos A\cos D, \quad Y = V\sin A\cos D, \quad Z = V\sin D. \tag{4}$$

To determine the coordinates of the radiant, we write out Airy's equations (see Eqs. (6) and (21), Chapter XIX)

$$-X\sin\alpha + Y\cos\alpha = kr\mu_\alpha\cos\delta, \tag{5}$$

$$-X\cos\alpha\sin\delta - Y\sin\alpha\sin\delta + Z\cos\delta = kr\mu_\delta, \tag{6}$$

where the sign on the left has been changed, since we are concerned with the velocity of the cluster relative to the Sun, and not the velocity of the Sun relative to the cluster. This equation can be written separately for each star in the cluster, and the internal motion of stars in the cluster will be treated as random errors in μ_α and μ_δ. The redundant equations (5) and (6) can be reduced by the least squares method to yield X, Y, and Z, which define the radiant of the cluster.

In practice, however, this method is inapplicable to most clusters. Open clusters generally occupy a small portion of the sky and the coefficients before the unknowns in Eqs. (5) and (6) therefore hardly change from one star to another. The determinant of the normal system will thus be close to zero, and the errors in the unknowns will be excessively large, so that X, Y, and Z will be physically meaningless. Since in this case the right-hand sides of the equations also differ insignificantly from one another, the redundant systems degenerate into two equations, (5) and (6), which only show that the radiant lies on the great circle defined by the direction of the average proper motion of the stars in the cluster.

Some open clusters, however, are large enough for the coefficients before the unknowns in Eqs. (5) and (6) to be substantially different, so that X, Y, and Z are determined with fair certainty.

In practice, since r is unknown, it is eliminated between (5) and (6) and following Charlier, we consider the redundant system

$$X(\mu_\delta\sin\alpha - \mu_\alpha\cos\delta\cos\alpha\sin\delta) - Y(\mu_\delta\cos\alpha + \mu_\alpha\cos\delta\sin\alpha\sin\delta) + Z\mu_\alpha\cos^2\delta = 0, \tag{7}$$

which gives the ratio $X:Y:Z$. This is sufficient for determining the coordinates of the radiant from the relations

$$\operatorname{tg} A = \frac{Y}{X}; \quad \operatorname{tg} D = \frac{Z}{\sqrt{X^2+Y^2}}.$$

Star clusters whose radiants can be determined are generally called moving clusters. Historically this term was adopted since some moving clusters were actually discovered in the result of a comparison of the proper motions of the constituent stars. Thus, the moving cluster containing the 13 stars of Ursa Major was discovered because five out of the seven brightest stars of this constellation have almost identical proper motions.

In fact moving clusters are open clusters which have large angular dimensions because of their proximity to the Sun. The large size is further accentuated by the low star density of some open clusters.

To improve the radiant coordinates, we should use the differential method, which amounts to the following. Let A_1 and D_1 be the approximate coordinates of the radiant , θ the position angle of the star's proper motion, $\Delta\theta$ the angle between the direction of the star's proper motion and the direction from the star to the radiant. Dividing (6) by (5), we get

$$-\frac{X\cos\alpha\sin\delta - Y\sin\alpha\sin\delta + Z\cos\delta}{-X\sin\alpha + Y\cos\alpha} = \frac{\mu_\alpha\cos\delta}{\mu_\delta} = \operatorname{ctg}\theta. \tag{8}$$

Differentiating and seeing that X, Y, and Z are expressed in terms of A_1 and D_1 by (4), we get

$$g\,\Delta A + h\,\Delta D = -\operatorname{cosec}^2\theta\,\Delta\theta. \tag{9}$$

Here g and h are known functions of α, δ, A_1, and D_1, which can be easily computed. In Eq. (9) the differentials have been replaced with error symbols, which is permissible if the errors are small.

Eq. (9) is written for each star. Since $\Delta\theta$ can be calculated, the redundant system (8) gives the corrections ΔA and ΔD to the radiant coordinates.

The radiant of a moving cluster can be similarly found from the radial velocities. To this end we use the equation

$$X\cos\alpha\cos\delta + Y\sin\alpha\cos\delta + Z\sin\delta = v_r, \tag{10}$$

which is generally applied to determine the apex of the Sun (see Eq. (24), Ch. XIX). The method gives the coordinates X, Y, and Z, and not the ratio $X:Y:Z$, so that we can further calculate the spatial velocity of the cluster.

If the X, Y, and Z obtained from the proper motions or the radial velocities are corrected for the spatial motion of the Sun, the resulting X', Y', and Z' give the radiant coordinates corrected for the motion of the Sun. This radiant is a point lying in the direction parallel to the residual spatial velocity of the center of inertia of the moving cluster.

Knowledge of the radiant (without correction for the Sun's motion) gives reliable parallaxes of individual stars in the moving cluster.

Indeed, Eq. (10) can be written in the form

$$V\cos\varrho = v_r,$$

where V is the spatial velocity of the star in the moving cluster relative to the Sun, and ϱ is the angle between the direction to the radiant and the

direction to the star. We need only the radial velocity of one star in the cluster in order to find the common V of the cluster. Then from the equality

$$V \sin \varrho = v_t = 4.74 \mu r$$

we can calculate the distance of the star

$$r = \frac{V \sin \varrho}{4.74 \mu}, \qquad (11)$$

if its proper motion is known. The so-called group parallaxes determined in this way are highly reliable. Their accuracy improves for near stars and for clusters which move with high velocities relative to the Sun.

Seven moving clusters are known at present. Three of these, Hyades, Pleiades, and Praesepe, are observed as visible open clusters: their actual position is marked by a substantial enhancement of surface star density. The other four moving clusters comprising stars from the constellations of Ursa Major, Perseus, Scorpius—Centaurus, and Coma Berenices, were discovered only because of pronounced likeness in proper motions of the stars. The star density in these clusters is exceedingly low. If we consider only stars which actually belong to the moving cluster, the density is seen to be less than the galactic field density. For example, the Ursa Major cluster contains only 13 stars in a volume of some 300 pc^3; the star density of the stars in this cluster is thus 1/3 of the star density near the Sun. In the other three moving clusters of this group the star density is even lower.

Thus, if we do not distinguish between moving and open clusters, we are led to the conclusion that many of the open clusters have exceedingly low star densities and may be detected only if they are near enough to the Sun, have large proper motions, and therefore give themselves away as moving clusters.

Sometimes similarity of proper motions of stars lying in widely different parts of the sky at greatly differing distances from the Sun is used as a justification for including these stars in a single moving cluster. Thus Roman (1949) and Ballario (1949) respectively assigned 135 and 290 stars to the moving cluster in UMa. The 135 on the former list, however, occupy an enormous volume no less than 500 pc in diameter, which covers the region occupied by the Sun, Pleiades, and Hyades. An elementary calculation shows that this volume should contain even more than 135 stars whose spatial velocities will show accidental likeness to the spatial velocities of the stars in UMa cluster.

§ 172. Stellar associations. While analyzing the distribution of O and early B stars over the sky, Ambartsumyan established that these stars form condensations in certain parts of the sky, which cannot be explained by random fluctuations of the apparent distribution. These condensations cannot be attributed to the distribution of dark matter in the Galaxy either. The hot O—B2 giants thus occur in separate groups, which Ambartsumyan called O associations.

The O associations were extensively studied at the Byurakan Observatory. The list published in 1952 by Markaryan contains 19 O associations and six probable O associations.

The sizes of these associations are between 40 and 200 pc. The number of O—B2 stars in these associations is between 10 and 100. The distance of the farthest known association is estimated at 2750 pc. They all, except the association in Orion, lie in the galactic plane. This is quite natural since the hot giants have a highly pronounced galactic concentration. The Orion association occupies the region of the sky between galactic latitudes $-10°$ and $-25°$.

O associations are apparently also observed in other spiral galaxies. The condensations visible in the spiral arms of the other galaxies are in fact groups of white-blue giants and supergiants.

Studies of the structure of O associations have shown that they generally contain open star clusters, which can be treated as the nucleus of the association. These clusters are made up of O—B2 stars, which often form an unstable configuration — star chains and Trapezium-type multiple stars, i.e., multiple systems which, according to Ambartsumyan's definition, contain at least three stars at comparable distances from one another. Numerous O associations are connected with gaseous nebulae. The volume occupied by O associations shows an increased proportion of B3—B9 stars as well. We do not know at this stage whether stars of later types also occur in increased density in associations.

The composition of the O associations is characterized by the color — apparent magnitude diagram (Figure 186) and the spectrum — absolute magnitude diagram (Figure 187) constructed by Blaauw and Hiltner (1959) for the Cepheus III association.

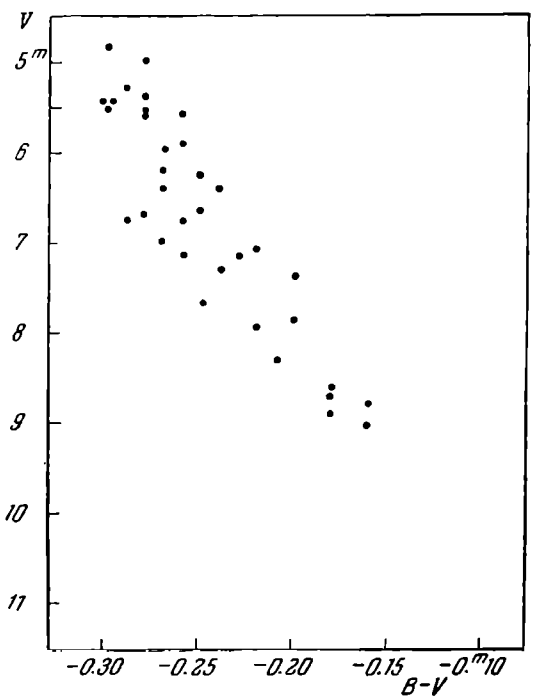


FIGURE 186

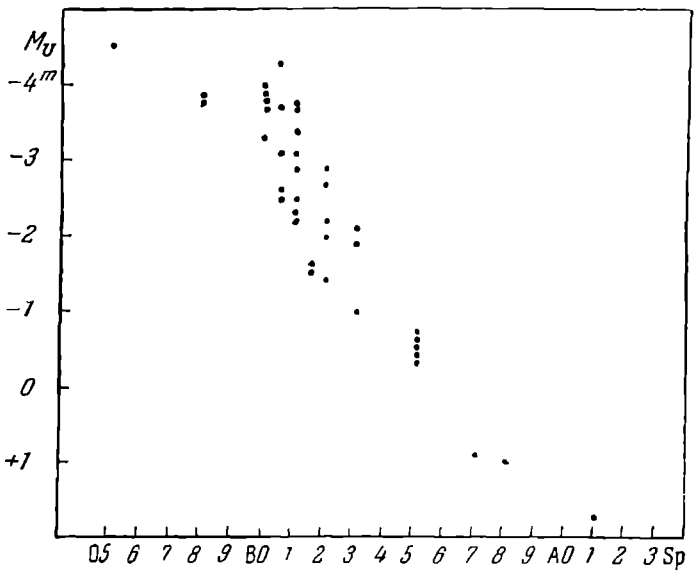


FIGURE 187

The density of matter in O associations seems to be lower than the density required for a stable formation. Hence follows the conclusion that O associations are inherently unstable. This conclusion is important because it shows that O associations must be young formations. Ambartsumyan advanced further arguments demonstrating that the O associations and the constituent stars are all relatively young objects: 1) O associations contain Wolf—Rayet stars, whose mass is expended at an extremely high rate and which therefore never reach old age; 2) normal O—B2 stars cannot maintain their usual rate of mass and energy consumption for a long time either; 3) O associations contain unstable multiple stars and star chains, again objects which cannot reach old age.

For an O association to be internally stable, i.e., to have negative energy, the velocities of the constituent stars relative to the center of inertia should be exceedingly small, of the order of 100 m/sec. This condition follows from the low density of matter in the O associations, which leads to a low critical velocity. Yet the dispersions of radial velocities and proper motions of the constituent stars in associations are quite considerable. Moreover, according to Blaauw, the proper motions of stars in the association Perseus II point to systematic radial expansion with a velocity of some 10 km/sec. A similar effect was observed by Markaryan for Cepheus II. Although the radial expansion of associations has not been established conclusively, since some observers failed to detect it (Woolley, Eggen, Lebedinskii), the overall instability of these formations leaves no doubt.

All these results led Ambartsumyan to the conclusion that O associations are the site of star formation, and the processes of star formation in O associations continue to this day.

The tendency to occur in physical groups is also observed for T Tau stars — irregular variables. These stars are late-type dwarfs or subgiants, with characteristic emission lines in their spectra. The groups of T Tau stars were named T associations by Ambartsumyan. They generally contain from 10 to less than a hundred T Tau stars. An exception to this rule is the T association in Orion, which contains 220 objects of this category. On the average T associations extend over several tens of parsecs. They often occur near dust nebulae. T associations show a pronounced galactic concentration, although it is not as marked as the concentration of O associations.

In 1950 Kholopov published a list of 21 T associations. The revised list published in 1959 contains 29 T associations and 12 possibles.

Since T associations consist of moderate luminosity stars, the known T associations are observed at distances not exceed 500 pc. It therefore follows that the actual number of T associations is much greater than the number of O associations.

The density of matter in T associations is low, and it therefore seems that these are young and unstable formations. The presence of emission lines in the spectra of T Tau stars is another argument in favor of the young age of these stars. Ambartsumyan is of the opinion that T associations, like O associations, are the site of star formation processes in the Galaxy.

§ 173. Globular clusters. Globular clusters are incomparably richer in stars than open clusters and they are characterized by regular spherical shape. The stellar population of globular clusters reaches tens of thousands, hundreds of thousands, and even millions of stars. Some clusters are not perfectly spherical, showing a certain ellipticity. The globular clusters are not sharply outlined. As in the case of open clusters, their boundary is best determined by star counts. Figure 188 is a photograph of a classical globular cluster M13 in Hercules.

Table 6 lists star counts to a given magnitude v in the globular cluster M92 made by Taylor (1954) on photographs taken with the 20-in. and the 100-in. telescope. The first column indicates the limits of the annular zones (in seconds of arc) in which the counts were made.

TABLE 6

R'' \ v	12.90	13.38	14.00	14.64	15.24	15.81	16.39	16.99	17.50	19.25
0″ – 21″.8	5	16	18	41	—	—	—	—	—	—
21.8 – 45.3	4	6	17	34	—	—	—	—	—	—
45.3 – 66.4	4	8	14	29	51	89	144	—	—	—
66.4 – 88.7	1	4	6	15	24	58	92	131	—	—
88.7 – 111.4	2	2	3	8	19	42	74	106	152	—
111.4 – 133.3	1	5	6	10	15	34	51	86	132	611
133.3 – 155.5	2	4	7	13	19	33	44	79	115	526
155.5 – 177.7	1	1	2	8	13	22	31	54	80	418
177.7 – 199.9	2	3	5	7	11	14	26	43	67	333
199.9 – 221.8	0	0	0	3	6	15	20	29	43	246
221.8 – 243.9	0	1	1	5	9	17	25	33	48	200
243.9 – 276.7	0	1	3	4	12	18	23	39	52	246
276.7 – 320.7	0	1	2	4	9	17	24	33	44	215
320.7 – 397.2	2	3	10	16	21	29	41	55	87	252

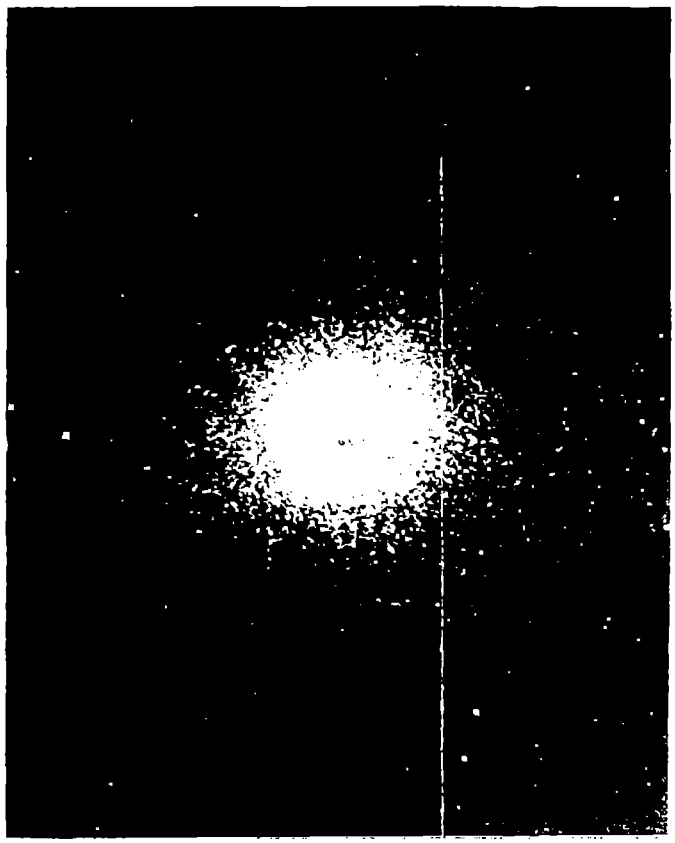


FIGURE 188

In the central rings the weak stars are so numerous that despite the high resolving power of the large telescopes the photographs show a continuous luminous background. Stars up to 15.24 mag. could not be counted in the first two central rings, and stars up to 19.25 mag. are not resolved in the five inner rings.

The star counts also include field stars which do not belong to the cluster. Their significance is particularly pronounced in the outer regions of the cluster, where the cluster star density is relatively low. The average number of field stars up to $v = 19^m.8$ and $v = 18^m.0$ per 1 sq. deg of the sky at the galactic latitude corresponding to M92 (approximately 33°) is 1.21 and 0.58, respectively. The results of Taylor before and after correcting for the field stars in each sq. deg at six different distances from the center of the cluster are listed in Table 7.

Table 7 shows that for stars up to $v = 18^m.0$ the radius of the cluster is approximately 550". For weaker stars with $18^m.0 < v < 19^m.8$ the boundary of the cluster lies somewhat farther: the average number of these stars in 1 sq. deg at a distance of 553".6 from the center (2.32 − 0.67 = 1.65) is still a factor of 2.5 greater than the average number of field stars (1.21 − 0.58 = 0.63) in the same area.

TABLE 7

R^*	Average number of stars to given v in 1 sq. deg			
	Without correction for field stars		Corrected for field stars	
	$v=19.8$	$v=18.0$	$v=19.8$	$v=18.0$
128".6	194.91	34.45	193.70	33.87
159 .8	131.95	20.74	130.74	20.16
223 .8	55.92	8.59	54.71	8.01
319 .8	19.59	3.11	18.38	2.53
448 .0	5.55	1.28	4.34	0.70
553 .6	2.32	0.67	1.11	0.09

Shapley's classification of globular clusters is based on the degree of concentration toward the center. Type I clusters are those with maximum central concentration, and type XII includes globular clusters with minimum central concentration. The assignment of individual clusters to different types is therefore based on subjective estimates, and not star counts. Some data indicate that globular clusters with higher concentration are characterized by larger diameters and greater luminosity.

The color index — apparent magnitude diagram of globular clusters is substantially different from that of the open clusters. It primarily shows a well developed branch of red giants and supergiants. At the same time the globular clusters contain absolutely no white-blue supergiants; B stars are either absent altogether or present in very small numbers.

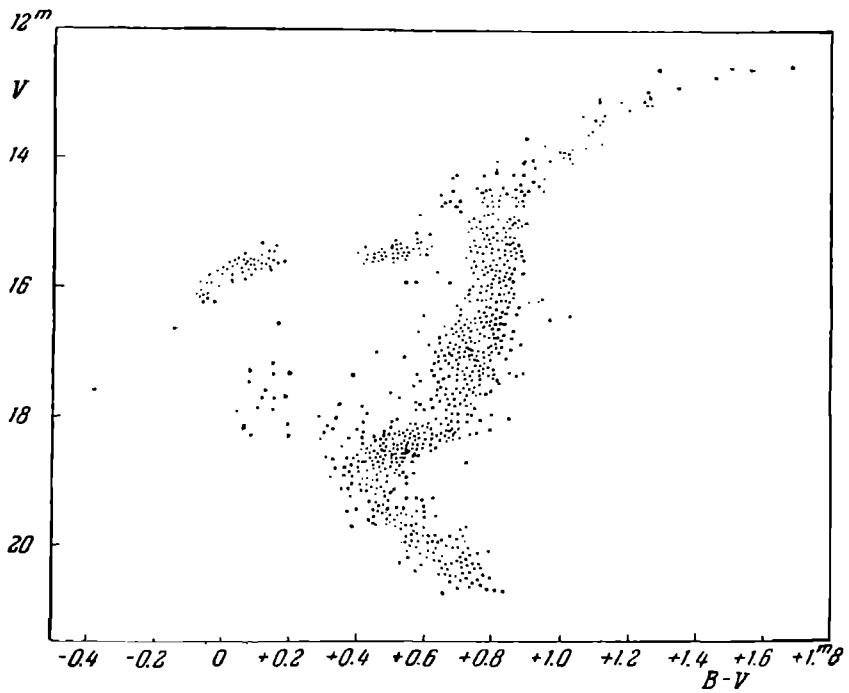


FIGURE 189

Figure 189 is the color index — apparent magnitude diagram of the globular cluster M3 plotted by Johnson and Sandage in 1956 using observations with the 200-in. and the 100-in. telescope. The diagram thus includes stars up to 21 mag. The main sequence in M3 is thus only partly represented, covering the range $B - V \geqslant +0^m.4$. White giants of luminosity class V are very few, and there are no white supergiants. The giant branch including exceedingly high luminosity stars is very prominent and rich. The so-called globular cluster sequence between the giant branch and the visible section of the main sequence is also densely populated. It creates two characteristic breaks in the diagram. Finally, there is a horizontal sequence of white giants which extends through the Hertzsprung—Russell gap; this sequence has a break for $B - V = 0^m.1 - 0^m.3$. Note, however, that the variables of this cluster (over 200) are not shown in the diagram, whereas 85 variables fall precisely in this break.

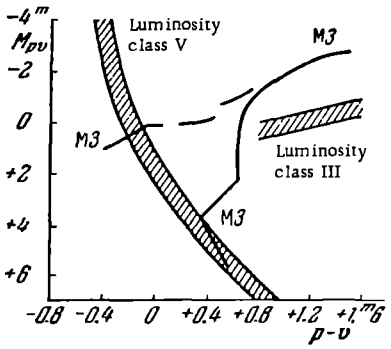


FIGURE 190

The difference between Population II stars (globular clusters) and Population I stars (Sun's neighborhood, open clusters) clearly emerges from Figure 190, where Sandage superimposed the color — luminosity diagrams of the globular cluster M3 and the luminosity sequences V and III, representative of Population I stars.

The color index—apparent magnitude diagrams of various globular clusters are on the whole similar, though not identical. For example, the horizontal branch in the globular cluster M92 is sparingly populated with white giants, and this cluster is much poorer in variables.

To determine the luminosity function of the stars in M3, Sandage carried out counts of 44,500 stars up to $m_p = 22^m.5$. The distance modulus of the cluster was estimated as $15^m.63$. The results are presented in Table 8.

TABLE 8

m_p	M_p	$N(m)$	m_p	M_p	$N(m)$
14.1	−1.53	169	19.5	+3.87	9812
15.0	−0.63	369	20.1	+4.47	15 940
16.1	+0.47	1318	20.7	+5.07	22 175
16.9	+1.27	1949	21.6	+5.97	34 875
17.7	+2.07	2648	22.5	+6.87	44 540

These data give the differential luminosity function $\varphi(M)$ of M3 (Figure 191). In this diagram, $\varphi(M)$ is the number of stars with absolute magnitudes between $-0^m.1$ and $+0^m.1$.

A bimodal curve was obtained, with maxima at $M = 0$ and $M = +5^m.5$. The first of the two maxima is significant according to Sandage; it is

attributed to the white stars of the horizontal sequence. The second maximum may be due to incompleteness of counts at high magnitudes. The luminosity function of M92 has no maximum at $M = 0$. This is explained by the much smaller population of the horizontal-sequence white stars in M92. Note that the luminosity function of stars in the plane of the Galaxy (Parenago) is also bimodal, but the maxima are shifted several stellar magnitudes in the low luminosity direction compared to the luminosity function of M3.

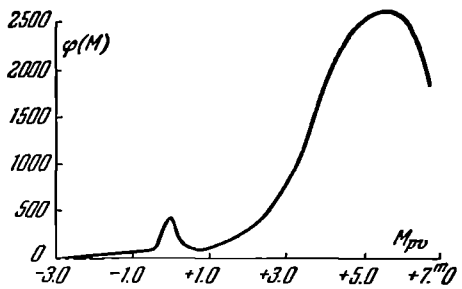


FIGURE 191

In globular clusters the mean luminosity of late-type stars is higher than that of the early-type stars; for example, the average apparent stellar magnitudes for stars of different spectral types in NGC 6205 (determined from the color index) are the following:

B	A	F	G	K
$15^m.20$	$15^m.01$	$14^m.85$	$14^m.28$	$12^m.74$

Even if we correct for the selective absorption of light, which reduces the luminosity of early type stars to a greater extent than the luminosity of late type stars, the late type stars retain their higher luminosities.

The spectra could be studied only for fairly few globular clusters. They were found to possess a number of peculiar features. Thus, the absorption in the CN λ 4215 Å line in G5—K2 stars in globular clusters is substantially weaker than in normal stars of the same spectral type and luminosity. A number of lines in spectra of other types also show increased or reduced intensity compared to the spectra of corresponding normal stars.

The globular clusters are exceedingly rich in variables. Variables of the following types were identified: 1) RR Lyrae stars (short-period cepheids), 2) W Vir stars, 3) RV Tau stars, 4) irregular and semi-regular variables. The RR Lyrae stars are the most frequent. There are indications that certain properties of variables in globular clusters differ from the corresponding properties of variables of the same type outside the clusters.

Alongside with globular clusters which contain very numerous variables, there are globular clusters with a small variable population. These clusters generally lie at the two extremes of the concentration scale, i.e., types I—II and XI—XII.

Globular clusters are spherical system with a high number of gravitating bodies. They can be compared to a certain extent to a gas sphere without any energy sources (if the radiation pressure is ignored). On the other hand, the density distribution in globular clusters can be obtained by direct star counts. Globular clusters thus can be used for verifying the basic postulates of thermodynamics and statistical physics.

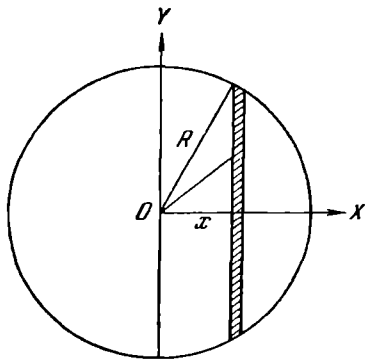


FIGURE 192

An important problem is the derivation of the true star density function of clusters from the observed surface density function. This problem was first formulated and solved by Zeipel in 1909. In 1911 Plummer advanced a simpler solution, which is described in the following.

Consider a rectangular coordinate system with its origin at the center of the cluster (Figure 192). The axes OX and OY lie in the plane of the sky, which coincides with the plane of the drawing. The axis OZ is directed along the line of sight. Let $D(r)$ be the spatial star density at distance r from the center of the cluster and $F(x)dx$ the number of visible stars in a strip of width dx distant x from the axis OY; R is the radius of the cluster. Evidently,

$$F(x)\,dx = dx \int\int D(r)\,dy\,dz, \tag{12}$$

where $r^2=x^2+y^2+z^2$, and the limits of integration are specified by the inequality $x^2+y^2+z^2 \leqslant R^2$.

Eq. (12) is a two-dimensional integral equation for the unknown function $D(r)$. It is readily inverted.

Changing over to the polar coordinates in the YOZ plane, we write $\xi^2 = y^2 + z^2$; $\varphi = \operatorname{tg}^{-1}\frac{z}{y}$. Then

$$F(x) = \int_0^{2\pi}\int_0^{\sqrt{R^2-x^2}} D(r)\,\xi\,d\xi\,d\varphi = 2\pi \int_0^{\sqrt{R^2-x^2}} \Phi(r)\,d\xi;$$

since $r^2 = x^2 + \xi^2$, changing back to the original integration variable we obtain

$$F(x) = 2\pi \int_x^R rD(r)\,dr.$$

The right-hand side of this equation is only a function of the lower limit of integration. Differentiating with respect to x, we obtain the sought function $D(x)$:

$$D(x) = -\frac{1}{2\pi x} F'(x). \tag{13}$$

Eq. (13) solves our problem. The function $F(x)$ is obtained from star counts in successive strips; the field stars must be subtracted from these counts.

In 1959 Oort applied Sandage's star counts to determine the spatial density of stars of various luminosities in the globular cluster M3. His results are listed in Table 9.

TABLE 9

M_v / r, pc	$< +3.5$	$+3.5 - +4.6$	$+4.6 - +6.3$
0	248	135	64
1	178	100	50
2	69.6	42.5	25.0
4	12.1	9.0	6.9
8	1.24	1.17	1.28
15	0.149	0.182	0.299
30	0.0125	0.0208	0.0387
50	0.0018	0.0030	0.0057
100	0.00013	0.00022	0.00040

We see that at the center the star density is exceptionally high, reaching many hundreds of stars per cubic parsec. Toward the periphery, the density falls off rapidly. The density of high-luminosity, i.e., high-mass stars, falls off at a faster rate.

In 1949 Kholopov considered an analogous problem for ellipsoidal globular clusters.

We can now compare the function $D(r)$ obtained from star counts with the theoretical model of a free gas sphere. Let p be the pressure at some point in the gas sphere, D the density, t the temperature, γ the ratio of specific heats at constant pressure and constant volume. Since the pressure at any point is proportional to density and temperature, we have

$$p = atD. \tag{14}$$

On the other hand, Poisson's adiabatic equation of state is

$$p = bD^{\gamma}. \tag{15}$$

The variation of pressure with distance is given by the equation

$$\frac{dp}{dr} = -Gm\frac{D}{r^2}, \tag{16}$$

where m is the mass inside a sphere of radius r. The change of mass is described by the equation

$$\frac{dm}{dr} = 4\pi r^2 D. \tag{17}$$

Eliminating p and m between Eqs. (15)—(17), we get

$$\frac{d}{dr}\left(\frac{br^2}{D}\frac{dD^{\gamma}}{dr}\right) + 4\pi G r^2 D = 0,$$

or, appropriately choosing units of length and density,

$$\frac{d}{dr}\left(\frac{r^2}{D}\frac{dD^{\gamma}}{dr}\right) + r^2 D = 0. \tag{18}$$

This is the so-called Emden equation, which is solved numerically. An analytical solution can be found only for $\gamma = 2$ and $\gamma = 1.2$. In the former case, the sought function has the form

$$D(r) = \frac{1}{r} \sin r,$$

which is physically meaningless, since the density cannot assume negative values.

The case $\gamma = 1.2$, which gives

$$D(r) = c_1 (1 + c_2 r^2)^{-\frac{5}{2}}, \tag{19}$$

is applicable to a gas of composite molecules, which corresponds to clusters comprising multiple stars.

For monoatomic and diatomic gases, we have $\gamma = 5/3$ and $7/5$, respectively.

Eq. (28) characterizes the so-called adiabatic state of a gas sphere, where temperature gradients are not leveled out by convection and other heat diffusion processes. In the other extreme case, when heat transfer processes virtually equalize the temperature throughout the sphere, t in Eq. (14) is constant, and we see that this equation corresponds to Eq. (15) with $\gamma = 1$.

Comparison of the true density function in globular clusters obtained from observations with the theoretical functions calculated from Eq. (18) has shown that (19) ensures a good fit only for the peripheral regions of some clusters. At the center, the star density increases toward the center at a faster rate than that predicted by Eq. (19). This can be interpreted as the result of an isothermal star gas. We are thus led to suggest the following model of globular clusters: the bulk consists of high-luminosity massive stars which should be compared with an isothermal gas sphere; this sphere is surrounded by an adiabatic "atmosphere" comprising relatively weak stars. However, since the multiplicity of stars and the isothermal conditions have similar effects — they both reduce the ratio γ — we can hardly distinguish between the two contributions. We should further remember that the theory and the calculations ignore the differences in stellar mass, an effect which is probably of considerable significance.

The distribution of color indices over the cluster radius reveals substantial reddening toward the center, which may be attributed to a high central concentration of the massive red supergiants. On the basis of this fact, Zeipel proposed a method for the determination of the relative masses of stars of various color indices.

Let the globular cluster be in a state of isothermal equlibrium. As we have observed in the preceding, star counts are compatible with this assumption everywhere, except in the peripheral parts. The Gibbs distribution function of elementary particles in the phase space is a function of energy of the form

$$\Phi = \alpha e^{-2hE}. \tag{20}$$

In expanded form, we may write

$$\Phi(u, v, w, r, m) = \alpha(m)\, e^{-hm[u^2+v^2+w^2-2P(r)]}, \tag{21}$$

where m is the stellar mass, u, v, w are the star velocity components, r is the distance from the center of the cluster, h is a constant dependent on the total energy (the "gravitational temperature") of the cluster, $P(r)$ is the potential energy, $\alpha(m)$ is a constant determined by the total number of stars of given mass.

Integration over all possible values of u, v, and w gives

$$\Phi(r, m) = \beta(m)\, e^{2hmP(r)}, \tag{22}$$

where $\beta(m)$ is a new function independent of r.

If the mass of a star of a certain type is taken as unity, we have

$$\Phi(r, 1) = \beta(1)\, e^{2hP(r)}. \tag{23}$$

Dividing (22) through by (23) raised to the power of m and extracting the logarithm, we obtain

$$\ln \Phi(r, m) = \ln \frac{\beta(m)}{\beta^m(1)} + m \ln \Phi(r, 1). \tag{24}$$

Taking counts of stars of two different types and then using Eq. (13) for conversion from apparent densities to spatial densities, we end up with known functions $\Phi(r, m)$ and $\Phi(r, 1)$. Inserting the values of these functions for various r in Eq. (24), we obtain a redundant system of equations which can be solved to give m.

The above method was applied by Zeipel, Lindgran, Wallenquist, and other authors to various globular clusters. The results on the whole are fairly consistent. They show that in globular clusters the masses of red giants are approximately a factor of 1.5 greater than the masses of the white stars, and the supergiant masses are double the masses of the white stars. This result is of considerable interest, since supergiants, being extremely rare, do not occur among binaries and their masses have never been determined.

Note, however, that the method of mass determination for globular clusters is far from being exact, since the conditions are never precisely isothermal and h is not quite constant, as assumed in our calculations. In fact h is a function of r. Moreover, the color indices of stars may acquire a biased distortion due to various photographic effects. Shapley, for instance, is of the opinion that the observed reddening of stars near the center is a result of the photographic Eberhardt effect, and not a significant structural feature of the cluster.

Mayall's spectra of 50 globular clusters give the following distribution according to spectral types:

A5—A9	6 clusters
F0—F4	7 clusters
F5—F9	13 clusters
G0—G5	24 clusters

The distances of globular clusters are determined by the following methods:

1. The method of cepheids, in its usual form.

2. The method of bright stars. It utilizes the fact that the difference between the apparent magnitudes of cepheids of a given period

and the mean apparent magnitude of the brightest stars in the cluster is almost constant for clusters containing cepheids. Hence follows the conclusion that the absolute magnitude of the brightest stars in globular clusters remains fairly constant in different clusters and can be used as a yardstick. According to Shapley, the average absolute magnitude of the 25 bright stars which come after the five brightest stars of a cluster (these are excluded, to avoid the effect of accidental projection onto the cluster) reveals a very slight dependence on the type of cluster. In type I clusters the average apparent magnitude of variables is $1^m.34$ greater than the average magnitude of the 25 brightest stars, and in type XII clusters the difference is $0^m.92$. The method of bright stars is applied to clusters which contain visible variables.

3. The method of diameters is based on the assumption that the linear diameters of globular clusters are virtually constant, or else that the diameters of clusters of one type are equal, so that Eq. (2) can be used. According to Parenago, Kukarkin, and Florya, the average linear diameter of globular clusters whose distances have been determined by the method of variables or the method of bright stars is 40 pc for type XII clusters, gradually increasing with the increase in central concentration and reaching 98 pc for type I clusters. A positive feature of this method is that it does not involve any absorption corrections.

4. The method of integrated brightness is used assuming a small dispersion of the absolute magnitudes of globular clusters. M_c — the average absolute magnitude of clusters of a given type — is determined using clusters of known distances. The distances of all the other clusters are then calculated from the distance modulus $m - M_c$. A correction for interstellar absorption should be introduced. According to Parenago, Kukarkin, and Florya, the average absolute integrated magnitudes of globular clusters increase from $-8^m.1$ for type I to $-5^m.4$ for type XII.

There are currently 118 known globular clusters. Some of these, although classified as galactic clusters, are very distant and are in fact extragalactic objects in their own right. In this category we have the globular cluster in Ursa Major, whose distance is estimated at 130 kpc (Shakhbazyan). Burbidge and Sandage also obtained 130 kpc for the distance of the globular cluster in Sextans.

The galactic concentration of globular clusters is much less pronounced than that of the open clusters, as is evident from a comparison of the data in the first two columns of Table 10 (which gives the frequency of directions and distances for globular clusters) and Table 4. The globular clusters form a halo subsystem.

Because of the skewed position of the Sun in the Galaxy, almost all the globular clusters lie between 240 and 40° longitude, i.e., in the hemisphere enclosing the center of the Galaxy.

Mayall (1956) and then Baade (1958) and Kinman (1959) established a definite relation between the physical characteristics of globular clusters and their spatial distribtuion. Clusters of the earliest integrated spectral type (A5—G0) have a low relative content of metals and are situated at comparatively large distances from the plane of the Galaxy, forming an extended halo with insignificant concentration toward the center of the Galaxy. Late-type clusters (G3—G5) with metal-rich stars show a relatively pronounced concentration toward the plane of the Galaxy. Clusters of integrated types G1—G2 occupy an intermediate position on this scale.

TABLE 10

Galactic latitude	Number of clusters	Galactic latitude	Number of clusters	Distance, kpc	Number of clusters
−90°— −30°	11	0°—40°	15	<4	6
−30°— −15°	15	40°—80°	2	4—8	18
−15°— −6°	24	80°—120°	1	8—12	8
−6°— −0°	9	120°—160°	5	12—16	13
0°— +6°	13	160°—200°	2	16—20	6
+6°— +15°	18	200°—240°	5	20—24	2
+15°— +30°	13	240°—280°	10	24—28	3
+30°— +90°	15	280°—320°	22	28—32	2
		320°—360°	56	>32	6
Total	118	Total	118	Total	64

Because of the tremendous distances to the globular clusters, their proper motions should be exceedingly small. Gamalei used measurements made at a large interval of time in order to determine the proper motions of seven globular clusters. Three clusters had proper motions of the order of 0".01, and the other four even less. Nevertheless, these minute proper motions are indicatory of tangential velocities much greater than 100 km/sec.

The radial velocities of globular clusters can be determined with fair reliability, and they have been measured so far for about 70 clusters. If the distance of the Sun from the center of the Galaxy and the circular velocity of the Sun are known, the eccentricities of the orbits of globular clusters in the Galaxy can be statistically estimated from their radial velocities. Gerner and then Kinman have shown that the globular clusters move in highly elongated orbits.

Table 11 lists the fundamental data for 12 globular clusters.

TABLE 11

NGC	Messier or other designation	l	b	R	r, kpc	m_p	Sp	v_r	T	m^*_{25}	Δm
104	47 Tuc	272°	−45°	23.0	5.8	3.0	G5	−24	III	13.44	0.35
4590	68	268	+37	2.9	11.5	7.6	A6	−116	X	14 80	0.42
5139	ω Cen	277	+15	25.0	5.0	3.6		+230	VIII		1.3
5272	3	7	+77	9.8	13.8	4.5	F4	−153	VI	14.23	0.26
5904	5	322	+46	12.7	8.3	3.6	F7	+49	V	13.97	0.35
6121	4	319	+15	14.0	2.3	5 2	F	+65	IX	13.11	1.9
6254	10	343	+22	8.2	5 0	5.4	G0	−71	VII	14.06	0 67
6266	62	321	+ 6	4.3	6.9	7.0	F6	−75	IV	15.87	2.4
6402	14	349	+13	3.0	7 2	7.4	G0	−129	VIII	15.44	2.9
6981	72	3	−34	2.0	18.2	8.6	G2	−255	IX	15.86	0.45
7078	15	33	−28	7.4	15.1	5.2	F0	−107	IV	14.31	0.53
7099	30	355	−48	5.7	12.6	6 4	A7	−174	V	14 63	0.34

In Table 11 T is Shapley's classification in terms of central concentration, m^*_{25} is the average apparent magnitude of the 25 brightest stars in each cluster, Δm is the absorption of light from the cluster in stellar magnitudes.

§ 174. Stability of star clusters. Star clusters are systems of gravitating bodies. The dynamics of star clusters is concerned with gravitational forces of four different classes: 1) the regular (general) field of the cluster,

2) the regular (general) field of the Galaxy, 3) forces due to random encounters between the stars in the cluster, 4) forces due to random encounters of the cluster stars with galactic-field stars or other objects.

The general field of the cluster is in fact the force which consolidates the cluster as such. It continuously changes the magnitude and the direction of the velocity vector of an individual star relative to the center of inertia of the cluster, and the star trajectory therefore remains confined within a small volume of space containing the cluster's center of inertia. The regular field of the cluster thus operates to retain the individual stars in the cluster.

The regular field of the Galaxy has a two-fold effect. First, the resultant force of this field applied to the center of inertia of the star cluster determines its orbit in the Galaxy. This motion clearly has only an indirect effect on the evolution of the cluster. Second, the regular galactic field applied to individual stars is different at different points throughout the cluster, and this gives rise to a tidal force which operates to tear the cluster apart. The tidal force increases as the cluster becomes larger and as the gradient of the galactic field is increased. Tidal forces are counteracted by the regular forces of the cluster. They increase with increasing mass and density of stars in the cluster.

The stability of the cluster in relation to the tidal forces produced by the galactic regular field was investigated by Bok, Mineur, and Chandrasekhar.

Let the center of inertia of the cluster lie in the plane of the Galaxy at a distance R_0 from the center. Suppose it moves in a circle with constant angular velocity ω. Consider a stationary rectangular system of coordinates where the axis Z is along the galactic rotation axis, and the axes X and Y lie in the plane of the Galaxy. We will also use a moving rectangular system of coordinates with its origin at the center of inertia of the cluster, with the axis Z_1 directed along the axis of galactic rotation and the axis X_1 invariably aimed to the center of the Galaxy. The moving system clearly rotates with angular velocity ω and the coordinates of points in the two systems are related by the equalities

$$\left.\begin{aligned} x &= (R_0 + x_1)\cos\omega t - y_1 \sin\omega t, \\ y &= (R_0 + x_1)\sin\omega t + y_1 \cos\omega t, \\ z &= Z_1, \end{aligned}\right\} \tag{25}$$

where the time t is reckoned from the time when the axes X and X_1 coincided.

Lagrange's function of a cluster star of mass m is

$$L = T - m\Phi - U', \tag{26}$$

where $m\Phi$ is the potential energy of the star in the galactic field; it is evaluated remembering that the distance of the star from the galactic rotation axis is

$$R = \sqrt{(R_0 + x_1)^2 + y_1^2}. \tag{27}$$

U' is the potential energy of the star in the cluster, expressed by

$$U' = mG \sum_i \frac{m_i}{r_i},$$

where r_i is the distance between the particular star and the i-th star of the cluster.

To find the kinetic energy of a star, T', we differentiate (25) with respect to t. Then, using the above equalities, we get

$$T' = \frac{m}{2}(\dot{x}^2+\dot{y}^2+\dot{z}^2) =$$
$$= m\left[\frac{1}{2}(\dot{x}_1^2+\dot{y}_1^2+\dot{z}_1^2) - \omega\dot{x}_1 y_1 + \omega\dot{y}_1(R_0+x_1) + \frac{1}{2}\omega^2(R_0+x_1)^2 + \frac{1}{2}\omega^2 y_1^2\right].$$

Lagrange's equations of the second kind

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} = \frac{\partial L}{\partial q_i} \quad (q_1 = x_1,\ q_2 = y_1,\ q_3 = z_1)$$

take the form

$$\left.\begin{aligned} \ddot{x}_1 - \omega\dot{y}_1 &= \omega\dot{y}_1 + \omega^2(R_0+x_1) - \frac{\partial\Phi}{\partial x_1} - \frac{1}{m}\frac{\partial U'}{\partial x_1}, \\ \ddot{y} + \omega\dot{x}_1 &= -\omega\dot{x}_1 + \omega^2 y_1 - \frac{\partial\Phi}{\partial y_1} - \frac{1}{m}\frac{\partial U_1}{\partial y_1}, \\ \ddot{z} &= -\frac{\partial\Phi}{\partial z_1} - \frac{1}{m}\frac{\partial U'}{\partial y_1}. \end{aligned}\right\} \tag{28}$$

From Eq. (27) we have

$$\frac{\partial\Phi}{\partial x_1} = \frac{R_0+x_1}{R}\frac{\partial\Phi}{\partial R}; \qquad \frac{\partial\Phi}{\partial y_1} = \frac{y_1}{R}\frac{\partial\Phi}{\partial R}; \qquad \frac{\partial\Phi}{\partial z_1} = \frac{\partial\Phi}{\partial z}. \tag{29}$$

The diameter of a cluster is small compared to R_0, and therefore dropping terms of second and higher order in x_1 and y_1, we write

$$R = R_0 + x_1;$$

Taylor expansion of Eqs. (29), putting $\left(\frac{\partial\Phi}{\partial z}\right)_0 = 0$ (the Galaxy is symmetric about the principal plane), gives

$$\left.\begin{aligned} \frac{\partial\Phi}{\partial x_1} &= \left(\frac{\partial\Phi}{\partial R}\right)_0 + x_1\left(\frac{\partial^2\Phi}{\partial R^2}\right)_0, \\ \frac{\partial\Phi}{\partial y_1} &= y_1\left(\frac{1}{R}\frac{\partial\Phi}{\partial R}\right)_0, \\ \frac{\partial\Phi}{\partial z_1} &= z_1\left(\frac{\partial^2\Phi}{\partial z^2}\right)_0, \end{aligned}\right\} \tag{30}$$

where the subscript indicates that the function should be evaluated at the center of inertia of the cluster.

Inserting (30) in (28) and seeing that $\omega^2 = \left(\frac{1}{R}\frac{\partial\Phi}{\partial R}\right)_0$, we get

$$\left.\begin{aligned} \ddot{x}_1 - 2\omega y_1 + \alpha_1 x_1 &= -\frac{1}{m}\frac{\partial U'}{\partial x_1}, \\ \ddot{y}_1 + 2\omega\dot{x}_1 &= -\frac{1}{m}\frac{\partial U'}{\partial y_1}, \\ \ddot{z} + \alpha_3 z_1 &= -\frac{1}{m}\frac{\partial U'}{\partial z_1}, \end{aligned}\right\} \tag{31}$$

where

$$\alpha_1 = \left(\frac{\partial^2\Phi}{\partial R^2} - \frac{1}{R}\frac{\partial\Phi}{\partial R}\right)_0; \quad \alpha_3 = \left(\frac{\partial^2\Phi}{\partial z^2}\right)_0.$$

To develop the right-hand sides of (31), we require the density distribution in the cluster. For simplicity, we assume that the smoothed-out density in the cluster is represented by a homogeneous triaxial ellipsoid. The potential energy then takes the form

$$U' = -\frac{1}{2}m\beta_0 + \frac{1}{2}\pi GmD(\beta_1 x_1^2 + \beta_2 y_1^2 + \beta_3 z_1^2),$$

where D is the density of the ellipsoid, the coefficients β_1, β_2, β_3 are positive, depending only on the semiaxis ratio $a : b : c$,* and β_0 depends additionally on D.

Equations of motion (31) thus take the form

$$\left.\begin{aligned} \ddot{x}_1 - 2\omega\dot{y}_1 + (\alpha_1 + \pi GD\beta_1)\,x_1 &= 0, \\ \ddot{y}_1 + 2\omega\dot{x}_1 + \pi GD\beta_2 y_1 &= 0, \\ \ddot{z}_1 + (\alpha_3 + \pi GD\beta_3)\,z_1 &= 0. \end{aligned}\right\} \tag{32}$$

Since $\alpha_3 > 0$ and $\beta_3 > 0$, integration of the third equation yields

$$z_1 = c_3 \cos\left[\sqrt{\alpha_3 + \pi GD\beta_3}\,(t - t_0)\right],$$

where c_3 and t are constants. The coordinate z_1 of all the stars thus remains bounded, and the cluster is therefore stable in the direction z_1.

In the first two equations in (32) we seek solutions in the form

$$x_1 = c_1 e^{ikt}; \quad y_1 = c_2 e^{ikt}, \tag{33}$$

where c_1, c_2, k are constants. Inserting these solutions in the corresponding equations, we get

$$\left.\begin{aligned} (\alpha_1 + \pi GD\beta_1 - k^2)\,c_1 - 2\omega ikc_2 &= 0, \\ 2\omega ikc_1 + (\pi GD\beta_2 - k^2)\,c_2 &= 0. \end{aligned}\right\} \tag{34}$$

The homogeneous system has a nontrivial solution only if its determinant vanishes. The resulting biquadratic equation for k is

$$k^4 - [\alpha_1 + \pi GD(\beta_1 + \beta_2) + 4\omega^2]\,k^2 + \pi GD\beta_2(\alpha_1 + \pi GD\beta_1) = 0,$$

and for

$$\alpha_1 + \pi GD\beta_1 > 0 \tag{35}$$

the two roots k^2 are positive and k are therefore all real. In this case, as we know, solutions (33) are bounded periodic functions. If now

$$\alpha_1 + \pi GD\beta_1 < 0, \tag{36}$$

solutions (33) contain hyperbolic functions, x_1 and y_1 increase indefinitely and the cluster becomes unstable.

* Detailed tables of these functions were published by Mineur (Ann. d'Astrophys., 2, No. 1:199—213. 1939).

Note that

$$\alpha_1 = \left(\frac{\partial^2\Phi}{\partial R^2} - \frac{1}{R}\,\frac{\partial\Phi}{\partial R}\right)_0 = \left[\frac{d}{dR}(R\omega^2)\right]_0 - \omega^2 = 2\left[R\,\frac{d\omega}{dR}\right]_0 \omega = 4A(B-A),$$

where A and B are Oort's coefficients. The stability condition (35) thus takes the form

$$D > D^* = \frac{4A(A-B)}{\pi G\beta_1} \tag{37}$$

and the stability of the cluster at a given point in the Galaxy thus depends only on the density and the flattening of the cluster. Taking $A = 0.020$ km/sec/pc, $B = 0.013$ km/sec/pc, we get

$$D^* = \frac{0.198}{\beta_1} \text{ solar masses/pc}^3.$$

Table 12 lists the values of β_1 and the critical density D^* for ellipsoids of revolution flattened in the direction of the z_1 axis.

TABLE 12

$\frac{c}{a}$	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1.0
β_1	0.278	0.499	0.677	0.824	0,946	1.048	1.211	1.333
D^*	0.71	0.40	0.29	0.24	0,21	0.19	0.16	0.15

We see that highly flattened ellipsoids are less stable than slightly flattened ellipsoids. The critical density in all cases is markedly less than the mean density of such open clusters as the Pleiades, Praesepe, h and χ Persei, where $D = 1-2$ solar masses/pc^3. The stability of these clusters is particularly high since they are not greatly flattened.

The effect of forces due to random encounters of stars in the cluster is considered in §164. For open clusters this factor is of great importance, since the relaxation times are small. The open clusters have passed through a substantial portion of their evolutionary path while undergoing steady dissipation. For globular clusters the effect of dissipation is not so pronounced, although the regular shape of the globular clusters and the distribution of stars of various masses in these clusters show that the lifetime of globular clusters is greater than the relaxation time.

Let us now consider the effect of forces due to random encounters of cluster stars with galactic field stars. The importance of these forces was first stressed by Rosseland.

The galactic field stars exchange kinetic energy with the cluster stars they encounter. Since the velocity of the field stars relative to the center of inertia of the cluster is greater than the velocity of the cluster stars, the latter gain kinetic energy in these encounters and the cluster expands. Individual stars may acquire supercritical velocities after a single encounter with a foreign star and escape from the cluster. This, however, does not lead to a compression of the remaining part of the cluster, as in the case of interaction between the stars of the cluster. Conversely, the

escape of a star from the cluster due to interaction with extraneous factors reduces the potential energy per star in the cluster, while the kinetic energy per star remains as before. The cluster should therefore expand.

Interaction of the cluster stars with galactic field stars thus leads to expansion of the cluster. We have seen in § 164, however, that random forces due to encounters between the stars in the cluster lead to gradual dissipation and contraction of the remaining part of the cluster. Which of the two effects prevails?

It can be shown that for typical open clusters, such as the Pleidaes, where the star density is relatively high, the contraction due to dissipation following internal encounters prevails over the expansion due to encounters with galactic field stars. In low-density clusters, such as the moving cluster in Ursa Major, the second effect predominates.

REFERENCES TO CHAPTER XXI

1. AMBARTSUMYAN, V.A. On the dynamics of star clusters.— Uchenye Zapiski, Leningrad University, 4, No.22. 1938.
2. AMBARTSUMYAN, V.A. Stellar associations and the formation of stars.— Izvestiya AN SSSR, phys. ser., 14, No.1. 1950.
3. BARKHATOVA, K.A. Light absorption and the diameters of open star clusters.— Astron. Zh., 25, No.4. 1949.
4. BARKHATOVA, K.A. Open star clusters.— Astron Zh., 27, No.3 1950.
5. GUREVICH, L.E. and B.Yu. LEVIN. Evolution of self-gravitating systems.— Doklady AN SSSR, 60, No.5. 1950.
6. MARKARYAN, B.E. Classification of open (galactic) star clusters.— Bull. Byurakan Observ., No.5. 1950; No.9. 1951.
7. MARKARYAN, B.E. A revised list of O associations.— Doklady AN Armenian SSR, 15. 1952.
8. PARENAGO, P.P., B.V. KUKARKIN, and N.F. FLORYA. The system of globular clusters.— Trudy, Shternberg Astronomical Institute, 16, No.113. 1949.
9. KHOLOPOV, P.N. A numerical method for determining the spatial density of stars in a spheroidal cluster.— Astron. Zh., 26, No.5. 1949.
10. KHOLOPOV, P.N. A revised list of T associations and their members.— Astron. Zh., 36;295. 1959.
11. KHOLOPOV, P.N. Ellipticity of globular clusters.— Astron. Zh., 27, No.4. 1952.
12. BROWN, A. On the determination of the convergent point of a moving cluster from proper motions.— ApJ, 112, No.2. 1950.
13. BRUGGENCATE, P., ten. Sternhaufen.— Berlin. 1927.
14. COLLINDER, P. On structural properties of open galactic clusters and their spatial distribution.
15. DREYER, J.E.L. A new general catalogue of nebulae and clusters of stars.— Mem. RAS, 49, Part I. 1888.
16. DREYER, J.E.L. Index catalogue of nebulae found in the years 1888—1894.— Mem. RAS, 51. 1895.
17. DREYER, J.E.L. Second Index Catalogue of nebulae and clusters of stars.— Mem. RAS, 59, Part II. 1908.
18. HALTON, C.A. Color-magnitude diagrams for seven globular clusters.— Astron. J., 60, No.9. 1955.
19. LOHMANN, W. Die Entfernungen der kugelformigen Sternhaufen.— Zeitschr. für Astrophys., 30, 1952.
20. MOWBRAY, A.G. The diameters of globular clusters.— ApJ, 104:1. 1946.
21. PLUMMER. The distribution of stars in globular clusters.— MN, 76, 1915.
22. SANDAGE, A.R. The luminosity function for the globular cluster M3.— Astron. J., 59:162. 1954.
23. SHAPLEY, H. Star Clusters.— New York. 1930.
24. TRUMPLER, R.J. Preliminary results on the distances, dimensions and space distribution of open star clusters.— Lick Observ. Bull., 14, No.420. 1930.
25. TRUMPLER, R.J. Galactic star clusters.— ApJ, 91, 1940.

T.A. AGEKYAN

Chapter XXII

EXTRAGALACTIC ASTRONOMY

§ 175. Introduction. Many of the nebulae seen in the sky are star systems which lie outside our Galaxy. In terms of their composition and size these star systems are similar to our Galaxy — the Milky Way — and are correspondingly called galaxies. Of the 103 objects in the Messier catalogue, 34 are independent galaxies. Their proportion in the NGC and IC (see §169) is even higher. A comprehensive catalogue of weak nebulous objects will contain but a negligible fraction of nebulae which are not galaxies.

The first extragalactic nebula was discovered and described as early as 1612 by Simon Marius (Germany). This was the sprial galaxy in Andromeda. The question whether the spiral and elliptic nebulae were galactic or extragalactic objects was finally settled in the 20th century. In 1917 Ritchey and Curtiss discovered novae in the nearest galaxies. In 1924—1926 Hubble resolved the outer regions of some spiral galaxies into individual stars through the 100-in. telescope, and in 1944 Baade resolved into stars the nuclei of these spiral galaxies and some elliptic galaxies. It was thus conclusively established that these objects were of extraglactic origin, island universes not unlike our Galaxy itself.

§ 176. Hubble's classification of galaxies. A systematic study of the extragalactic objects was begun in the 1920's by Hubble. These years saw the advent of a new astronomical science — extraterrestrial astronomy.

In 1926 Hubble proposed the first classification of galaxies.

The galaxies are divided into regular and irregular. The regular galaxies are subdivided into spiral (S) and elliptical (E).

Spiral galaxies (often briefly called spirals) are highly flattened systems. Spiral galaxies observed in plan show a bright nucleus with spreading spiral arms. A spiral galaxy viewed edge on appears as a highly flatened ellipse with a central bulge (the nucleus) and sharp pointed edges. A dark band or lane extends along the major axis of the ellipse: it is produced by the absorbing dark matter which is concentrated in these systems near the plane of symmetry, as in our Galaxy.

Hubble divided the spiral galaxies into three subclasses, according to the relative size of the nucleus and the prominence of the spiral arms. His classification reflects the observational fact that the relative size of the nucleus shrinks as the spiral arms become more developed.

Spiral galaxies with poorly developed, sometimes rudimentary spiral arms and a large nucleus are designated Sa (Figure 193, NGC 2681, Sa, kS1; here and in what follows, the name of the galaxy is followed by

Hubble's classification and sometimes by Morgan's as well). Galaxies of this type viewed edge on are seen to be less flattened than other spirals (Figure 194, NGC 4594, Sa, kS6).

FIGURE 193

FIGURE 194

Spiral galaxies Sb have somewhat better developed spiral arms (Figure 195, NGC 2841, Sb) and their nuclei are smaller than in Sa galaxies.

Sc galaxies have highly developed spiral arms (Figure 196, NGC 5364, Sc, fg S4); in some cases the arms are separated to such an extent that the spiral structure appears to be decaying (Figure 197, NGC 598, Sc, fS3). Sc galaxies observed edge on are extremely flattened (Figure 198, NGC 4565, Sc).

In accordance with Jeans's hypothesis, which stipulates that the spiral galaxies evolve from Sa through Sb to Sc, the Sa spirals are called early, the Sb intermediate, and the Sc late.

FIGURE 195

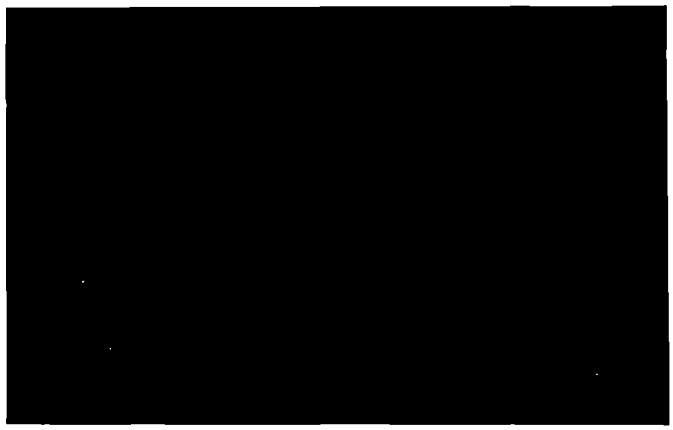


FIGURE 196

In some spiral galaxies the nucleus is situated at the middle of a "bar," and the spiral arms stretch from the two ends of this bar (Figure 199, NGC 2859, SBa; Figure 200, NGC 1300, SBb, fB2). These barred spirals receive an additional letter B in their classification, and by analogy with normal spirals they are designated SBa, SBb, SBc. These galaxies as if constitute a parallel branch to the Sa, Sb, Sc spirals.

The elliptical galaxies E, unlike the spiral galaxies, are featureless. Their surface brightness gradually falls off from the center toward the edges (Figure 201, NGC 221, E2; Figure 202, NGC 3115, E7).

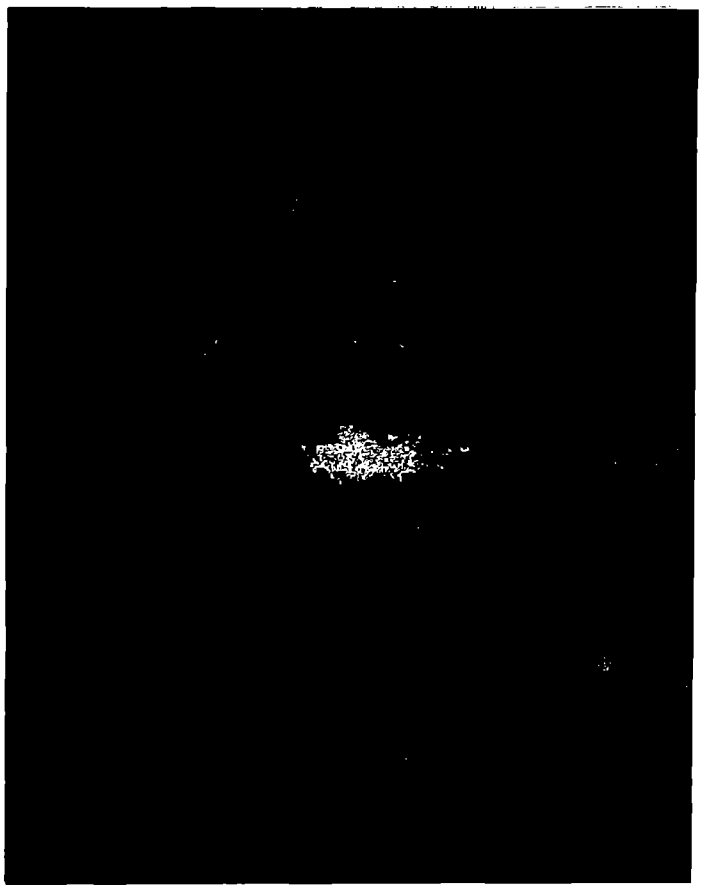


FIGURE 197

FIGURE 198

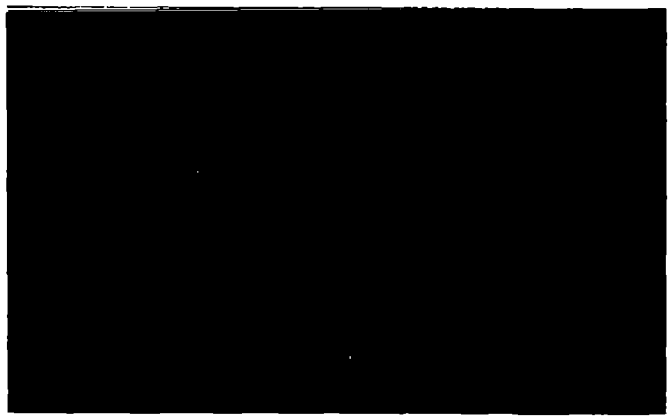


FIGURE 199

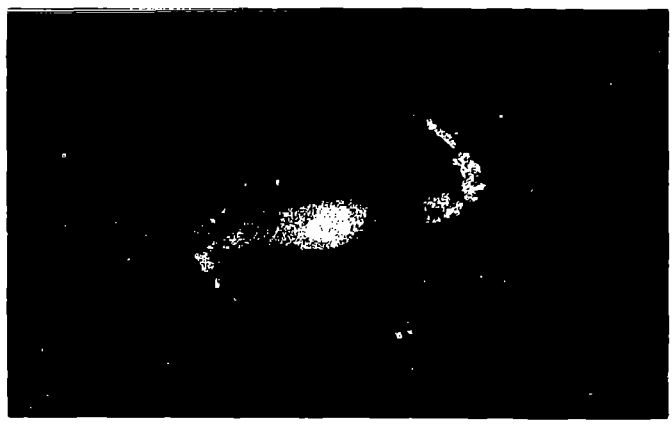


FIGURE 200

An additional characteristic of these galaxies is the flattening of the apparent ellipse, defined as

$$10 \cdot (1 - \eta), \tag{1}$$

where η is the sphericity of the apparent ellipse, equal to the length of the semiminor axis in units of the semimajor axis. The flattening of individual galaxies is generally expressed in whole numbers. Elliptical galaxies of maximum flattening are assigned degree of flattening 7, and are thus designated E7. Round elliptical galaxies are designated E0. All the intermediate flattening values are also observed. The flattening of the spiral galaxies viewed edge on is 8, 9, and 10. The apparent flattening of

a spiral, however, is not used as a physical characteristic of the galaxy, since it is mainly determined by the orientation of the system relative to the observer.

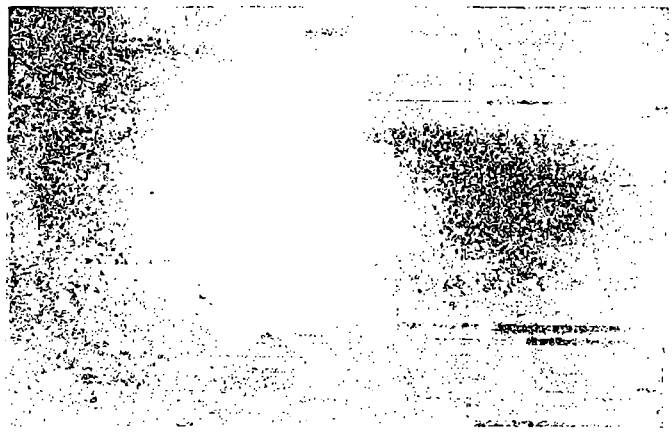


FIGURE 201

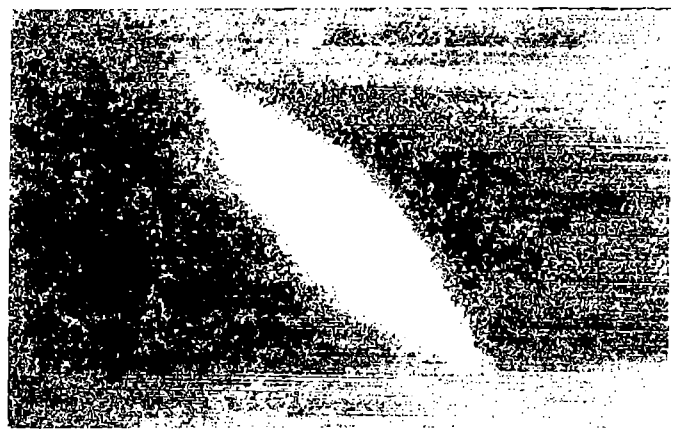


FIGURE 202

Irregular galaxies have no distinct symmetry (Figure 203, NGC 4449, II). They are designated by the letter I. The irregular galaxies were subsequently divided into two subtypes, which markedly differ from one another. Subtype I (Magellanic Clouds) is characterized by traces of spiral structure and fairly high luminosity. It is further distinguished by prevalence of Population I stars. Galaxies of this subtype are designated II. Subtype II, which includes the nearby galaxies in Sculptor and Fornax, has a highly peculiar distinctive feature: the luminosity is exceedingly low and no central concentration is observed. These galaxies show a trace of a nucleus, which is not found in elliptical galaxies. At the same time S0 galaxies have no spiral arms, and properly speaking are not spirals at all.

FIGURE 203

Schematically, Hubble's classification can be written in the following form:

$$E1 - E2 - E3 - E4 - E5 - E6 - E7 \begin{cases} Sa - Sb - Sc \\ SBa - SBb - SBc \end{cases} I.$$

This scheme may reflect to a certain extent the actual evolution of the galaxies. We can by no means assert, however, that Hubble's scheme is intended as a model of galactic evolution.

Hubble's classification has currently found wide uses. However, recent studies of a great number of galaxies have revealed enormous variety, especially among the spiral galaxies. Observations disclosed characteristic features not accounted for by Hubble's classification. For example, many of the galaxies classified as spiral actually have no spiral arms: they are ring-shaped (Figure 204, NGC 4826). Different intermediate forms were observed, and it was therefore realized that Hubble's classification required further elaboration. Thus, Vaucoulers (1957) proposed the following more detailed classification. He divided the galaxies into four principal types: elliptical (E), lenticular (S0), spiral (S), irregular (I). Lenticular and spiral galaxies fall into two subfamilies: ordinary (SA) and barred (SB), with an intermediate SAB type. Spiral galaxies occur in two varieties: spiral proper (s) and ring-shaped (r), with an intermediate rs type. Depending on the prominence of the spiral structure, Vaucouleurs distinguishes between early (a), intermediate (b), late (c), and very late (d) galaxies, and galaxies of the Magellanic Clouds type (Sm). The Magellanic Clouds in Hubble's classification are irregular galaxies. Vaucouleurs, however, detected traces of spiral structure in the two clouds and moreover pointed out a number of other galaxies of similar type. All these were double galaxies, and Vaucouleurs is of the opinion that we are facing in these cases the results of a substantial decay of spiral structure due to interaction between the two components.

FIGURE 204

Irregular galaxies II are thus classified by Vaucouleurs as Sm because of the observed traces of spiral structure. The irregular galaxies III are the only representatives of irregular galaxies in Vaucouleurs's classification. They are appropriately designated by the single letter I.

§ 177. Distances of galaxies. Extragalactic astronomy commonly operates with two large distance units: kiloparsec (kpc), equal to 1000 pc, and magaparsec (Mpc), equal to one million parsecs. Kiloparsecs are used in measuring the sizes of galaxies, the distances of the nearest galaxies, and the diameters of small clusters of galaxies. The distances to the far galaxies and clusters of galaxies and the diameters of large clusters are expressed in megaparsecs.

Correct determination of the extragalactic distances is one of the fundamental problems of extragalactic astronomy. The commonly used methods were mainly developed by Hubble.

1. The method of cepheids. The period of a cepheid is indicatory of its absolute magnitude. If the apparent magnitude is corrected for absorption, we can calculate the distance modulus. The method of cepheids

is clearly the fundamental yardstick of extraterrestrial astronomy and it gives highly accurate results. Until quite recently, this method was virtually irreplaceable for the calibration of all the other distance measuring techniques in extragalactic astronomy. In this respect, the situation is analogous to the role of trigonometric parallaxes for the calibration of all the other methods used in measuring galactic distances. Biased errors in the method of cepheids are therefore automatically carried over to all the other methods of extragalactic distance measurement. The method of cepheids in its turn is calibrated using trigonometric parallaxes, so that it is in fact this old traditional technique which constitutes the basis for all the methods of distance measurement in astronomy.

Only long-period cepheids were detected in other galaxies and could be used for distance measurement. The short-period cepheids (RR Lyrae stars) are intrinsically too faint to be resolved even with the largest instruments in the nearest galaxies.

2. The method of novae makes use of the fact that at the maximum the novae reach absolute magnitudes of about -7^m. When applying this method, we should take care to measure the apparent magnitude of the nova just near the maximum. Normally it is very difficult to detect the exact maximum, since the premaximum phase is characterized by a very rapid growth of brightness whereas the postmaximum phase involves a slow reduction of luminosity. As a result, the measurements give too low magnitudes and the distances are exaggerated. Reliable corrections for the mean delay can be introduced if a sufficient number of novae is observed in the particular galaxy.

Another method is to measure the magnitude of the nova a fixed number of days (10, 15, 20) after the maximum. At these times in the postmaximum phase the magnitude of the nova is -6, -5, -4^m, respectively; a few days after the maximum the decreasing branch is considerably more level than right after the maximum, and the error in pinpointing the exact maximum is not as critical as before. The main shortcoming of this method is that it is applicable only to the nearest galaxies, where the novae remain visible for a lengthy period after the maximum.

3. The method of bright stars is based on the assumption that the absolute magnitudes of the brightest stars are constant, at least in galaxies of one type. In particular, it is assumed that the brightest stars in spiral galaxies have the same luminosity as the brightest stars in our Galaxy. The analysis is geneally based on the mean apparent luminosity of some ten or twenty of the brightest stars.

4. The method of globular clusters, like the above method, is based on the assumption that the globular clusters in other galaxies on the average have the same absolute magnitude as the globular clusters in our Galaxy.

5. The method of ring diameters was recently proposed by Vaucouleurs (1957), who found that the diameters of the rings in ring-shaped galaxies are approximately equal for the different subtypes (Sra, Srb, Src, Srd), whereas the average diameter of a Sc and Sd spiral is double that of a Sa spiral. The average ring diameter is 2.5 kpc, with a mean deviation of 0.4 kpc. If the linear diameter of the ring is known, the distance can be inferred from its angular size. The largest telescopes will measure rings $0'.1$ in diameter. This corresponds to distances of 80 Mpc,

which are definitely inaccessible to any of the above methods. The advantage of this method is that it does not require any correction for light absorption. The method of ring diameters has been hardly used in practice, however.

6. The method of integrated magnitudes fixes the distance by comparing the apparent integrated magnitude and the absolute integrated magnitude, which is assumed constant for galaxies of a given type. The main significance of this method is that it will estimate the distances to the farthest galaxies, in which the individual objects required by the other five methods are no longer resolved. This method is inaccurate, since the dispersion of absolute magnitudes is considerable even for galaxies of one type.

7. The method of average diameters is based on the assumption that the linear diameters of galaxies of one type are equal, so that the angular size of the major diameter of a galaxy is only a function of its distance. This method is also inaccurate since the linear diameters show a considerable spread for each particular type. Moreover, the apparent angular diameter that we measure depends on the minimum surface brightness to which the detector responds.

8. The method of radial velocities. This method uses Hubble's red shift law: the lines in the spectra of galaxies show a red shift which is proportional on the average to the distance of the galaxy. If the red shift is attributed to the Doppler effect, we have the relation

$$v = H \cdot r, \qquad (2)$$

where v is the recession velocity of the galaxy, r is the distance, H is Hubble's constant.

According to Hubble's original results (1936), $H = 540$ km/sec/Mpc.

If the red shift is associated with some other factor, and not with the Doppler effect, Eq. (2) has no physical meaning and velocity v is to be understood symbolically. This, however, need not affect the application of Eq. (2) to distance dtermination.

Eq. (2) is satisfied only on the average. Each galaxy has its own peculiar velocity. The average peculiar velocities of the galaxies are estimated at 400—500 km/sec and are apparently little sensitive to the distance. Comparison of this figure with H shows that the method of radial velocities is valid for individual galaxies only at distances greater than 10 Mpc. For very distant objects, this method is highly reliable provided that the Hubble constant is accurately known.

§ 178. Apparent magnitudes, absolute magnitudes, diameters, and surface brightness of galaxies. Table 1 is a list of the 20 brightest galaxies published by Shapley in 1958.

The table shows that only three galaxies are visible to the naked eye: the Large Magellanic Cloud (LMC), the Small Magellanic Cloud (SMC), and the Andormeda Nebula, NGC 224 (M 31). There are six galaxies on the list with apparent magnitudes of up to $7^m.9$, six galaxies between $8^m.0$ and $8^m.9$, and seven between $9^m.0$ and $9^m.5$. This shows that the number of galaxies of a given apparent magnitude increases with increasing apparent magnitude. The brightest galaxies are equally distributed between the Southern and the Northern Hemisphere. Their distribution over the sky is distinctly

nonuniform, however. They almost completely avoid the zone $-20° < b < +20°$. This zone, which accounts for more than 1/3 of the total sky area, contains a single galaxy, against the expected seven. This is apparently associated with strong absorption of light at low latitudes. The light absorption near the galactic equator probably obscures several of the brightest near galaxies.

TABLE 1

No	Name or NGC	Messier	m_p	α	δ	l	b	Hubble's type	Morgan's type	r	Angular dimensions
1	LMC	—	1.2	5^h26^m	−69°	247°	−33°	I I		22	180′
2	SMC	—	2.8	0 50	−73	268	−45	I I		25	180′
3	224	M31	4.5	0 40.0	+41°00′	90	−20	Sb	kS5	230	200′×90′
4	598	M33	6.7	1 31.1	+30 24	102	−30	Sc	fS3	220	10′×10′
5	253	—	7.6	0 45.1	−25 34	91	−37	Sc	afS6		
6	55	—	7.8	0 12.5	−39 30	183	−84	Sc	aI		
7	5236	M83	8.0	13 43.3	−29 37	287	+32	Sc	fgS1	900	10′×8′
8	3031	M81	8.1	9 51.5	+69 18	109	+42	Sb	kS4	770	16′×10′
9	4594	M104	8.6	12 37.3	−11 21	269	+51	Sb	kS6p	2500	7′×1′,5
10	5457	M101	8.6	14 1.4	+54 35	66	+60	Sb	fS1	900	22′×22′
11	4826	M64	8.7	12 54.3	+21 47	246	+82	Sb	fgS4p	900	8′×4′
12	System in Sculptor	—	8.9	0 58.9	−33 50	226	−83	I II		70	45′
13	4736	M94	9.0	12 48.6	+41 23	83	+76	Sb	gS3	900	5′× 3′
14	System in Fornax	—	9.1	2 39.4	−34 34	201	−64	I II		150	50′
15	3034	—	9.2	9 51.9	+69 56	108	+42	I I	aI	500	13′×8′
16	4945	—	9.2	13 2.4	−49 01	273	+12	S			11′×2′
17	221	M32	9.3	0 40	+40 36	90	−22	E2	kE3	230	12′×8′
18	2403	—	9.3	7 32	+65 43	118	+30	Sc	aS4	500	10′×6′
19	5194	M51	9.3	13 27.8	+47 27	66	+68	Sc	fS1	900	14′×10′
20	205	—	9.5	0 37.6	+41 25	89	−21	E5	Ep	230	26′×16′

Table 2 is the list of the 34 nearest galaxies (including our Galaxy), and it was compiled from Holmberg's 1950 data. This list is clearly far from being complete. It should be supplemented with additional nearby but faint (dwarf) galaxies. Specifically note, that the M101 group contains no galaxies weaker than $12^m.7$. The list is apparently selective, giving preference to the brighter galaxies. Furthermore, a substantial number of near galaxies in low galactic latitudes is invisible because of interstellar absorption.

Using the data on the nearest galaxies, Holmberg obtained the average absolute magnitudes and dispersions for the different galaxy types listed in Table 3.

Hubble obtained for bright galaxies of all types $\bar{M}_p = -14^m.2$ and $\sigma_M = 0^m.85$.

This divergence is quite understandable since examination of the bright galaxies leads to selection of high-luminosity objects, increasing their average luminosity and reducing the dispersion. If weak near galaxies are detected and taken into consideration, the average luminosity of all types will decrease and the dispersion will somewhat increase. Nevertheless, the dispersion of absolute magnitudes of galaxies is definitely much less than the dispersion in the absolute magnitudes of stars, where $\sigma_M \approx 4^m$. Because of this difference in the dispersion of absolute magnitudes

TABLE 2

No.	Name or designation	Hubble's type	Morgan's type	m_p	M_p	Distance modulus		v, not corrected for the Sun's motion, km/sec	Color indices	Angular diameter	Linear dimension, kpc
						observed	corrected for absorption				
					The Local Group						
1	Galaxy	Sb and Sc									26
2	LMC	II		1.2	−15.9	17.1	16.7	+ 276		780'	4.6
3	SMC	II		2.8	−14.5	17.3	17.0	+ 168		180'	3.5
4	System in Sculptor	III		8.8	−10.6	19.4	19.2			45'	0.9
5	System in Fornax	III		9.1	−11.9	21.0	20.8	+ 149		50'	2.1
6	NGC 6822	I	L	9.1	−12.4	21.6	21.0	− 34		20'	0.9
7	NGC 147	E3	L	10.5	−11.9	22.4	21.5		+ 0.73	18' x 12'	1.0 x 0.7
8	NGC 185	E1	Ep	10.2	−12.2	22.4	21.5	− 270	+ 0.74	14' x 12'	0.8 x 0.7
9	NGC 205	E5	Ep	8.9	−13.5	22.4	21.8	− 239	+ 0.71	26' x 16'	1.7 x 1.0
10	NGC 221	E2	kE3	9.1	−13.3	22.4	21.8	− 220	+ 0.90	12' x 18'	0.8 x 0.5
11	NGC 224	Sb	kS5	4.3	−18.1	22.4	21.8	− 267	+ 0.86	197' x 92'	13.1 − 6.0
12	NGC 1613	I		10.0	−12.0	22.0	21.8	− 235	+ 0.39	23' x 23'	1.5 − 1.5
13	NGC 598	Sc	fS3	6.2	−16.1	22.3	21.9	− 190	+ 0.40	83' x 53'	5.8 x 3.7
14	Wolf - Lundmark system	III		11.1	−11.2	22.3	22.0			13'	1.0
15	NGC 6946	Sc	afS1	11.1				+ 34			
16	IC 10	Sc						− 343			
17	IC 342	Sc						− 20			
18	System in Leo	III									
19	System in Sextans	III									

					M81 Group						
20	NGC 2366	I	aL	12.6	−11.4	24.0	23.6	+ 140		10'	1.5
21	NGC 2403	Sc	aS4	8.8	−15.2	24.0	23.6	+ 66		23' x 15'	4.4 x 2.2
22	NGC 2976	Sc		10.7	−13.3	24.0	23.6	+ 40	+ 0.64	10' x 6'	1.5 x 9.0
23	NGC 3031 (M81)	Sb	kS4	7.8	−16.2	24.0	23.6	− 55	+ 0.85	35' x 14'	5.3 x 2.1
24	NGC 3034	I	aI	9.2	−14.8	24.0	23.6	+ 263	+ 0.81	13' x 8'	2.0 x 1.3
25	NGC 3077	I	Ep	10.6	−13.4	24.0	23.6	−130	+ 0.68	9' x 8'	1.4 x 1.2
26	NGC 2574	I		10.9	−13.1	24.0	23.6	+ 17	+ 0.29	16' x 8'	2.4 x 1.2
27	HO I	I		13.3	−10.7	24.0	23.6		+ 0.29	5' x 5'	0.8 x 0.8
28	HO II	I		11.1	−12.9	24.0	23.6		+ 0.34	10' x 10'	1.5 x 1.5
					M101 Group						
29	NGC 5194	Sc	fS1	8.9	−15.1	24.0	23.7	+ 438	+ 0.53	14' x 10'	2.2 x 1.5
30	NGC 5195	I	Ip	10.5	−13.5	24.0	23.7	+ 542	+ 0.98	9' x 7'	1.4 x 1.1
31	NGC 5204	Sc	fS4	11.2	−12.8	24.0	23.7	+ 270		10'	1.6
32	NGC 5457 (M101)	Sc	fS1	8.2	−15.8	24.0	23.7	+ 285	+ 0.30	28'	4.5
33	NGC 5474	Sc	fS1p	11.2	−12.8	24.0	23.7	+ 230	+ 0.31	7'	1.1
34	NGC 5585	Sc	fgS4	11.3	−12.7	24.0	23.7	+ 300	+ 0.35	9' x 6'	14 x 1.0

of stars and galaxies, 13 out of the 20 brightest galaxies are among the 33 nearest objects, while of the 20 brightest stars, only 4 are among the 33 nearest objects (Tables 6 and 7, Chapter XVIII).

TABLE 3

Type	Number of galaxies	$\overline{M}_p$	σ_M
E	6	−12,2	1,1
S	10	−14,8	1,8
I	12	−13,6	1,7
Total	28	−13,5	1,85

Using the data of Table 3, Holmberg constructed the luminosity function of the nearest galaxies. This function is shown in Figure 205 (curve II). The same figure also gives the luminosity functions of Hubble (I) and Zwicky (III). Hubble constructed his luminosity function for all the visible galaxies, i.e., all the galaxies up to a given apparent magnitude, assuming a normal luminosity distribution. Holmberg's luminosity function is skewed, with its mode at $-12^m.2$. Holmberg's function is closer to the actual luminosity function than Hubble's function, since the latter is distorted by observational selection. Note, however, that Holmberg's function is based on a very small sample, which includes only the nearest galaxies. Moreover, remember that the composition of the large clusters of galaxies is greatly different from the composition of the field galaxies (this is evident already from the differences in the frequency of flattening) and Holmberg's function may be considered only as the luminosity function of galaxies which are not included in any large cluster.

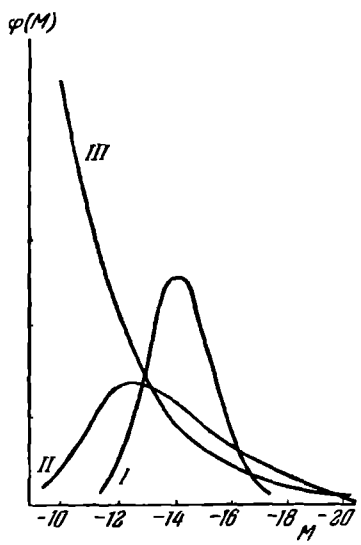


FIGURE 205

Some astronomers are of the opinion that the luminosity function has no maximum, that the number of galaxies increases monotonically with decreasing luminosity, and the observed maxima are the outcome of artificial selection, which virtually cuts off all weak galaxies from the observational samples. There are also indications of a gradual transition from galaxies to globular clusters and other star clusters of smaller scale. Thus, Zwicky (1957) proposed for the luminosity function of galaxies a monotonic function of the form

$$\varphi(M) = A\left[10^{0.2\cdot(M-M_0)} - 1\right], \qquad (3)$$

where the parameters should be determined using two points on the right branch of Holmberg's function, which corresponds to the distribution of high-luminosity galaxies. The function (3) clearly cannot be normalized. The corresponding curve (III) is also shown in Figure 205.

Baade opposed this idea of no clear-cut boundary between galaxies and globular clusters. He pointed out that the known galaxies of the lowest absolute luminosity are still 2—3 magnitudes brighter than the globular clusters of highest luminosity.

Finally note that Hubble's and Holmberg's luminosity functions and all the data in Tables 1 and 3 pertaining to distance moduli, absolute magnitudes, and linear dimensions should be radically revised: Baade (1953) showed that the scale of extraglactic distances based on Hubble's techniques was wrong (see § 183 for a more detailed discussion). All the extragalactic distance moduli have to be increased at least by $1^m.5$. The absolute magnitudes are thus reduced by the same amount. The average absolute magnitude of galaxies according to Holmberg (taking a correction of $1^m.5$) will drop to $-15^m.5$, the absolute magnitude of NGC 224 (M31) will be $-19^m.6$, which is equal to the absolute magnitude of the Galaxy (-19^m or -20^m). Our Galaxy is thus no longer an object of extraordinary luminosity, although it remains among the giant galaxies.

Since

$$r = 10^{0.2(m-M)} \qquad (4)$$

the increase of the distance modulus by $1^m.5$ increases the distance r by a factor of $10^{0.3}$ or almost double. All the linear dimensions of galaxies in Table 2 (except those of our Galaxy) should therefore be doubled. The size of NGC 224 becomes comparable with the size of our Galaxy, but the dimensions of other galaxies nevertheless remain substantially smaller. This significant difference is primarily due to the fact that the Galaxy and NGC 224 are really giant systems, and all the other nearby galaxies are substantially smaller. However, methods of higher sensitivity will clearly reduce the difference to a considerable extent.

The measured angular dimensions of galaxies are largely dependent on the particular method of measurement. If ordinary photographic techniques are used, the size of the galaxy increases with increasing speed, increasing exposure, increasing sensitivity of the plate, etc. Photoelectric techniques record farther lying outer regions of the galaxies. Finally, the star count method, which can be applied to the nearest galaxies resolved into individual stars, is apparently the only one which gives the true limits of the galaxies. Here the boundary is determined as the line beyond which the number of stars to a given magnitude per unit sky area shows no excess compared to the average number of field stars. Vaucouleurs (1955) applied the method of star counts to the Magellanic Clouds and obtained 1200' for the angular size of the Large Cloud and 540' for the Small Cloud. The angular dimensions of other galaxies are underestimated even to a greater extent in Table 3 (with the exception of NGC 224, which has been studied most comprehensively). It seems that Table 3 gives only the size of the inner, relatively bright parts of the galaxies. If an observer situated in another star system were to apply the same photographic methods to our Galaxy, the Sun would be left outside the Galaxy.

The large differences in the diameters of the galaxies in Table 3 show that the method of distance determination from the angular dimensions of galaxies is indeed inaccurate.

If we know the integrated apparent magnitude m of a galaxy and its angular dimensions α and β (expressed in minutes of arc), the average surface brightness of the galaxy in stellar magnitudes per sq. minute is $\overline{m}' = m + 2.51 \lg(\pi\alpha\beta)$. This magnitude is not a single-valued characteristic, since the angular dimensions of the galaxy depend on the particular measurement methods applied. The average magnitudes of the galaxies in some fixed system of measurement can nevertheless be compared. Thus Vaucouleurs, using the Shapley—Ames catalogue, obtained the following average surface magnitudes and mean square deviations from the average:

Type	$\overline{m}'$	σ
E0, S0	11.75	0.90
Sa, Sb, Sc	13.15	0.75

Thus, the average surface brightness of early-type galaxies is substantially less than that of late-type galaxies.

Hubble, Redman, Oort, and many others studied the distribution of surface brightness in galaxies. All elliptical galaxies were found to have identical surface brightness distribution. The only difference is the degree of flattening of the galaxies and the actual value of the surface brightness at the center of the ellipse. This result is of considerable cosmogonic significance, since it points to a common evolutionary path for all the elliptical galaxies.

Hubble (1930) proposed the following expression for the distribution of surface brightness along the major and the minor diameters of galaxies:

$$B = \frac{B_0}{(r + a)^2}, \qquad (5)$$

where r is the distance from the center, B_0 and a are the distribution parameters.

Vaucoulers's function provides a somewhat better approximation to the observed distribution of surface brightness in E galaxies:

$$\lg \frac{B}{B_0} = -3.33 \left(\sqrt{\frac{r}{r_e}} - 1 \right), \qquad (6)$$

where B_0 is the surface brightness at the center, r_e is some effective radius.

Spiral galaxies have no common brightness distribution function because of considerable differences in their external appearance. Shchegolev (1955) published a curve (Figure 206) of the surface brightness along the diameter of the spiral galaxy NGC 5457, viewed in plan. The solid curve gives the surface brightness in blue light, and the dashed curve that in yellow light; the units are apparent magnitudes per sq. second. A sharp peak is observed at the center of the galaxy. Smaller peaks flanking the nucleus on both sides correspond to spiral arms and smaller formations branching off the main arms. In blue light all the peaks associated with the spiral arms are sharper than those in yellow light. This shows that in spiral arms white-blue stars and gaseous nebulae are concentrated in greater numbers than the yellow and red stars. However, even in spiral arms, the radiation in yellow light is hardly smaller (or only slightly smaller) than the radiation in blue light.

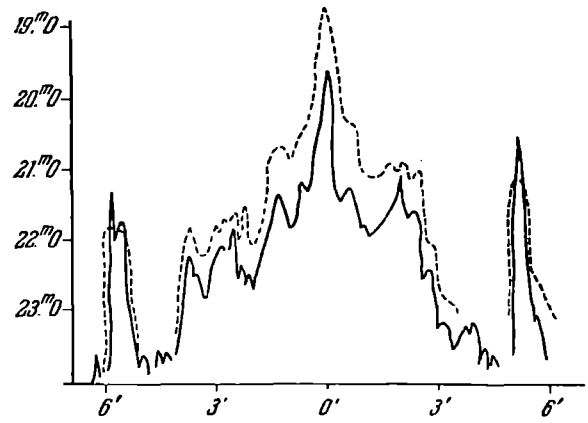


FIGURE 206

§ 179. **Spectra, color indices, and stellar populations of galaxies. Morgan's classification.** The spectra of galaxies are a superposition of the spectra of the constituent objects. These are generally spectra with absorption lines. If the objects with emission lines (mainly gas nebulae) make a substantial contribution to the total radiation of the galaxy, the resultant spectrum will also show emission lines.

The spectra of galaxies cannot be identified with the spectrum of any individual star because of their composite origin. However, comparing the intensities of various lines in the spectra of galaxies we can determine their "spectral type" in accordance with the methods of ordinary spectral classification of stars. The spectral type obtained in this way is evidently close to the weighted average (over the luminosity) spectral type of the constituent objects.

Table 4 lists the frequency of spectral classes among 572 galaxies of various types (Humason, Mayall, and Sandage, 1956).

TABLE 4

Spectrum \ Type		Sc + SBc	Sb + SBb	Sa + SBa	S0 + SB0	E
Emission	9	11				
A5—A6	1	2	1			
A7—A8		2	1	1	1	
A9—F0	1	12	5	1	2	1
F1—F2	1	4	3			
F3—F4		1	5		1	
F5—F6		12	9	6	5	1
F7—F8		10	9	8	3	4
F9—G0	1	5	25	10	11	10
G1—G2	1	11	10	18	25	33
G3—G4		5	22	18	39	54
G5—G6		1	5	19	23	54
G7—G8			7	3	7	21
G9—K0						

The mean error in the determination of the spectral type of a galaxy was estimated as approximately half a type. Therefore, for elliptical galaxies, the dispersion of spectral types in Table 4 is probably mainly due to errors in spectrum identification. For spirals, on the other hand, and mainly those of subtypes Sc + SBc and Sb + SBb, substantial dispersion of spectral types within a given subtype of galactic classification is obvious.

We further see from the table that earlier-type galaxies on the average have an earlier spectrum. Table 5 lists the average spectra of galaxies of each type (all emission spectra and irregular galaxies were omitted).

TABLE 5

Type	Number of galaxies	Average spectrum
E	178	G 3.7
S0 + SB0	117	G 2.2
Sa + SBa	84	G 1.4
Sb + SBb	102	F 9.6
Sc + SBc	65	F 6.1

The color index of galaxies increases on passing from early to late spectra. It should thus also increase on passing from early-type to late-type galaxies. This is confirmed by the data on the average color index for the different types of galaxies, obtained by Holmberg for 23 of the nearest galaxies, and by Stebbins and Whitford for 187 galaxies using the photoelectric technique (Table 6).

TABLE 6

Type	Holmberg		Stebbins and Whitford	
	N	*CI*	*N*	*CI*
I I	4	+0.33		
I II	3	+0.82		
Sc	8	+0.49	26	+0.47 ± 0.08
Sb	3	+0.84	37	+0.82 ± 0.12
Sa	1	+0.81	17	+0.82 ± 0.12
S0			32	+0.83 ± 0.14
E	4	+0.77	54	+0.86 ± 0.06

Holmberg's data show that irregular galaxies fall into two distinct groups in terms of their color index. I I have a smaller color index than Sc, whereas the average color index of I II is the same as for early-type spirals and E galaxies.

A highly complex problem is how to establish from the spectrum of a galaxy the presence and the numerical composition of some restricted group of objects, e.g., stars of a certain spectral type and given luminosity. In some cases, even this problem is solvable. Thus, the presence of distinct CN absorption bands in the spectrum reveals the presence of normal giants of spectral types G8—K3 in the galaxy, since only these stars are capable of producing these bands.

More general considerations — determination of the relative number of Population I or Population II stars — are much less certain. E and S0 galaxies seem to be composed of Population II stars, Sa and Sb galaxies are mixed Population I and Population II, the percentage of Population I stars increasing from Sa to Sb. Sc and II galaxies are the richest in Population I stars. In some systems Population I stars prevail, as in the Large Magellanic Cloud.

Resolution of the nearest galaxies into individual stars and careful examination of their stellar composition led to identical results.

The spectra of galaxies with a higher proportion of Population I stars are expected to show emission lines more frequently. This conclusion is confirmed by Table 7, which lists the data of Humason, Mayall, and Sandage for the percentage of galaxies of different types showing the emission line λ 3727 Å [O II].

TABLE 7

Type	Number of galaxies	Percentage of galaxies with emission lines
Sc + SBc	41	85
Sb + SBb	66	80
Sa + SBa	37	62
S0 + S00	52	48
E	82	18

The spectra of Sc and II also show lines of higher ionization potentials.

The color indices of elliptical galaxies are approximately constant over the entire surface of each galaxy. In spiral galaxies, the nucleus and the spiral arms have widely different color indices. The nucleus is much redder than the arms. According to Seyfert (1940), the average color index of the nucleus for seven spirals is $0^m.8$, and at certain points of the spiral arms it reaches $0^m.0$. The nuclei of spiral galaxies thus have the same color as the elliptical galaxies. They also have the same spectrum. Resolution of the nuclei of the nearest spiral galaxies into individual stars has shown that they are composed of Population II stars. The spiral arms contain an abundance of blue giants and supergiants, long-period cepheids, and gas and dust clouds, i.e., typical Population I environment.

When spiral galaxies are photographed through a red filter, the spiral branches are obliterated and the galaxy shows as a smooth featureless ellipse. This confirms the conclusion that the spiral galaxies are white-blue in color, but nevertheless contain a substantial number of yellow-red Population II stars.

In 1956 Haro reported the discovery of galaxies showing strong ultraviolet radiation. In some of these galaxies, the ratio of radiation densities in yellow, blue, and ultraviolet light is comparable with the corresponding ratio for stars of extreme violet color, e.g., white dwarfs or nuclei of planetary nebulae. The light of these galaxies shows a pronounced central concentration, so that the ultraviolet emission virtually comes from the nucleus. Their spectra show, besides the 3727 Å line, also lines of higher excitation potentials. Galaxies of different morphological types are represented in this group (there are definitely some spiral

and elliptical galaxies). Note that these violet galaxies are localized only in certain groups, in particular they are numerous in the Coma Berenices cluster.

Hubble's morphological classification was extended by Morgan and Mayall, who used the spectral characteristics of galaxies to incorporate the element of stellar population composition into the classification scheme. The absorption and emisssion lines of the spectra, despite their composite appearance, give reliable information on the composition of stellar population in galaxies. The spectral types (see Table 4) lie between A5 and G8, and although the spectrum shows a definite correlation with Hubble's types, this correlation is not particularly pronounced and in some types, especially Sc and Sb, the dispersion of the spectra is very considerable. Morgan and Mayall found that the correlation between the spectrum and the degree of central concentration is much more pronunced that the correlation between the spectrum and the type of the galaxy, or the correlation between the type and the central concentration. The later the spectral type, the higher is the central concentration. By specifying the degree of central concentration in galaxies we thus fix the spectral type and hence the stellar composition.

In Morgan's classification (1958) irregular galaxies and spiral galaxies without marked concentration toward the center are designated a. The violet part of the spectrum of these galaxies is attributed to the radiation of stars of spectral types B, A, and F. At the other end of the sequence lie the elliptical galaxies and the spiral galaxies with maximum central concentration. The population of these galaxies is designated by the letter k. The violet region in the spectra of the central parts of these galaxies is attributable to the emission of K-type giants. All the intermediate cases are obtained by interpolation in terms of the observed concentration between the two extreme cases.

The intermediate populations are designated af, f, fg, g, gk. Although the classification is based on the central concentration of radiation, the violet part of the spectrum is always attributable to the emission of the giants of the particular spectral type which is identified with the letter used for the classification.

The second parameter in Morgan's classification is the Hubble shape parameter: S for spirals, B for barred spirals, E for elliptical galaxies, I for irregular galaxies. The different subtypes are Ep for elliptical galaxies with a substantial content of dark matter, D for systems with pronounced axial symmetry without signs of spiral or elliptical structure. L for systems with low surface brightness (corresponding to type III), N for systems with a small bright nucleus against a weak background. The spiral galaxies are divided into subtypes Sa, Sb, Sc.

Systems with individual peculiarities receive an additional letter p.

The third parameter is the apparent flattening of the system. For spirals it is clearly the result of inclination to the line of sight, so that this parameter can be identified with inclination. Galaxies with the largest observed flattening are designated by the numeral 7, and spherical galaxies by 1. Nought is not used, to avoid confusion with Hubble's type S0. All the intermediate degrees of flattening are designated from 6 to 2.

Thus, NGC 224, which in Hubble's classification is a Sb galaxy, is designated kS5 in Morgan's classification, since this system has a very

pronounced concentration of radiation toward the center (the bulk of the galaxy is composed of Population II stars) and the galactic plane is markedly inclined to the tangent plane, producing substantial apparent flattening.

The galaxies shown in Figures 193, 194, 196, 197, and 200 are identified both according to Hubble's and Morgan's classification (see §176).

Morgan classified 605 galaxies in the new scheme. Morgan's classification is a major breakthrough in extragalactic astronomy. Note however, that the violet galaxies discovered by Haro do not satisfy the principles laid down by Morgan. These galaxies have a considerable central concentration, so that according to Morgan they should be composed of population k. In fact, population a of white-blue supergiants predominates in these galaxies, their nuclei included.

§ 180. **The frequency of galactic types.** In 1936 Hubble classified 600 bright galaxies and obtained the following percentage distribution:

E	Sa + SBa	Sb + SBb	Sc + SBc
17	19	26	38

Irregular galaxies were very few, and Hubble did not include them in his statistics.

The observed frequency of galactic types will coincide with the true distribution only if the average absolute magnitudes and their dispersion are identical for galaxies of all types. Otherwide, the apparent distribution will exaggerate the proportion of high-luminosity galaxies, since they are visible throughout a large volume of space, whereas the apparent number of low-luminosity galaxies will be too low as only the nearest of these are visible.

According to Hubble, the majority of the observed galaxies are spirals. From Table 3 we see, however, that spirals are on the average $2^m.6$ brighter (in terms of absolute luminosity) than the elliptical galaxies. The true relation of the different types is therefore entirely different from that predicted by the observed distribution. In 1948 Efremov showed that elliptical galaxies are in fact much more numerous than the spiral galaxies.

The abundance of irregular galaxies requires an even more radical revision. Statistical samples contained exceedingly few irregular galaxies and Hubble actually ignored them. The data in Table 3 show, however, that the I galaxies are of higher luminosity than the E galaxies and it would thus seem at a first glance that they are indeed fewer in number than E. This conclusion, however, is based on the peculiar composition of the class I in Table 3: it includes both the I I galaxies (i.e., the irregular MO types) and I II. Yet the absolute magnitude of I I is about $-15^m.5$, whereas the I II are much weaker. According to Wilson (1955), I II objects are visible only near the Galaxy. There are only six objects of this type with absolute magnitudes between -12^m and $-8^m.5$. The detection difficulties are further enhanced by their exceedingly low surface brightness. All the known I II galaxies were therefore detected in the result of the apparent concentration of the resolved stars. Some of the nearest I II galaxies have probably escaped detection to this day. If we take this point into consideration, we conclude that I II galaxies are most abundant in the immediate neighborhood of the Galaxy, which clearly suggests that the overall abundance of I II is

also high. This suggestion is confirmed by Zwicky's discovery (1957) of a great number of weak irregular galaxies with low surface brightness in the Virgo cluster. Because of their low surface brightness these galaxies had remained undetected until Zwicky photographed them with the high-speed 48-in. Schmidt at the Mount Palomar Observatory.

Note that the frequency of galactic types is clearly nonuniform in space; for example, around our Galaxy, small elliptical galaxies are represented in great numbers, and there are no giant elliptical galaxies. The Virgo cluster is rich in giant elliptical galaxies and contains spirals of all types. The Corona Borealis cluster contains hardly any Sa, Sb, and Sc spirals, whereas S0 are observed in great numbers.

In parallel with different stellar populations, we can thus speak of different pooulations of higher-order objects (clusters of galaxies).

§181. **Distribution of the flattening of galaxies.** The observed galactic shapes and dynamic considerations suggest that to first approximation regular galaxies are flattened ellipsoids of revolutions. This approximation provides a better fit to elliptical galaxies than to spirals.

Let the semimajor and the semiminor axes of an oblate ellipsoid of revolution be 1 and ξ. ξ is also called the sphericity or the true sphericity of the ellipsoid. The true sphericity is equal to the apparent sphericity η (i.e., the sphericity of the observed ellipse) only if the angle i between the ellipsoid axis of revolution and the plane of the sky is zero. For other i, we have $\eta > \xi$.

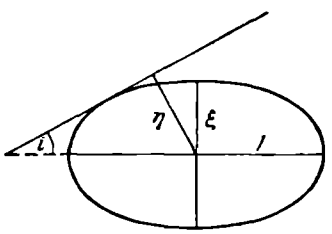


FIGURE 207

We require a relation among η, ξ, and i. In Figure 207, draw a tangent at an angle i to the ellipse with axes 1 and ξ. The equation of this tangent is

$$y = kx + \sqrt{\xi^2 + k^2}, \tag{7}$$

where

$$k = \operatorname{tg} i. \tag{8}$$

The distance of this tangent from the origin is

$$\eta = \sqrt{\frac{\xi^2 + k^2}{1 + k^2}}. \tag{9}$$

By (8) and (9),

$$\cos^2 i = \frac{1 - \eta^2}{1 - \xi^2}. \tag{10}$$

The apparent sphericity function $\varphi(\eta)$ is obtained from observations. We are interested, however, in the true sphericity function $f(\xi)$. To derive an equation relating these functions, we assume that all the possible directions of the galactic symmetry axis are equiprobable (this assumption is justified by a number of studies). The distribution function of i is then given by

$$\theta(i)\,di = \cos i\,di.$$

The probability that the two independent variables i and ξ fall between $[i,\ i+di]$, $[\xi,\ \xi+d\xi]$ is clearly $f(\xi)d\xi\cos i di$. Expressing i from (10) and integrating over all possible ξ, we obtain the probability that η lies in the interval $[\eta,\ \eta+d\eta]$:

$$\varphi(\eta)\,d\eta = d\eta\int_0^{\eta}\frac{\eta f(\xi)\,d\xi}{\sqrt{(1-\xi^2)(\eta^2-\xi^2)}}. \tag{11}$$

We thus have an integral equation for $f(\xi)$. Introducing new variables

$$h=\xi^2,$$
$$g=\eta^2,$$

we define the functions

$$F(h)=\frac{f(\sqrt{h})}{2\sqrt{h-h^2}},$$
$$\Phi(g)=\frac{\varphi(\sqrt{g})}{\sqrt{g}}.$$

Eq. (11) then takes the form

$$\Phi(g)=\int_0^{g}\frac{F(h)}{\sqrt{g-h}}\,dh. \tag{12}$$

Eq. (12), the well-known Abel equation, has the following solution:

$$F(h)=\frac{1}{\pi}\frac{d}{dh}\int_0^{h}\frac{\Phi(g)}{\sqrt{h-g}}\,dg.$$

Reverting to old variables and functions, we have

$$f(\xi)=\frac{2}{\pi}\sqrt{1-\xi^2}\,\frac{d}{d\xi}\int_0^{\xi}\frac{\varphi(\eta)\,d\eta}{\sqrt{\xi^2-\eta^2}}. \tag{13}$$

Eq. (13) enables us to convert the apparent sphericity function to the true sphericity function of galaxies.

This problem was first solved by Hubble (1926), who used a numerical method.

Kavraiskaya (1958) applied Eq. (13) to determine the true sphericity function for the field galaxies and the galaxies in clusters. For the field galaxies the true sphericity was found to fall between 0 and 0.4 and also between 0.95 and 1. The first of the two intervals (corresponding to true flattening of 9 and 10) clearly contains spirals, and the second interval (with true flattening of 0 and 1) contains elliptical galaxies of spherical and

almost spherical shapes. Cluster galaxies, on the other hand, have true sphericities between 0.5 and 0.9, i.e., the clusters are rich in elliptical galaxies with true flattening between 2 and 7.

Note that the apparent flattening of a galaxy is not a well-defined quantity. Galaxies do not have sharp boundaries and the length of each diameter of the apparent ellipse depends on the limiting surface brightness to which the particular measurement method responds, i.e., the speed of the optics, exposure time, sensitivity of the photoemulsion, etc. The isophotes of elliptical galaxies are ellipses of different sphericity, which definitely increases for the weaker isophotes.

It is therefore desirable to use observational series obtained with identical instruments. It is moreover always desirable to explicitly cite the minimum surface brightness used in the reduction of the galactic sphericities.

FIGURE 208

Our assumption that E and S galaxies are oblate ellipsoids of revolution to first approximation is clearly inapplicable to some star systems. Ogorodnikov has shown that, theoretically, there may exist elongated star systems corresponding to Jacobi's triaxial ellipsoids in the theory of stable configurations of rotating fluid masses. Some observational data indeed point in favor of the existence of such systems. The weightiest argument is clearly the case of barred spirals. Since we know that the principal plane of NGC 1300 passes through the spiral arms, we see from Figure 200 that the bar of this galaxy approximately has the shape of a prolate ellipsoid. On the other hand, there are galaxies which have a distinct bar with traces of rudimentary, extremely short spiral arms extending from the two ends (Figure 208). In this case the bar is naturally again a prolate ellipsoid. It now remains to make the last step in our reasoning: some of the highly flattened systems are in fact bars without any traces of spiral arms at the ends.

§ 182. The red shift in galactic spectra. In 1929 Hubble reported the discovery of the red shift: the lines in the spectra of galaxies, except the nearest ones, were shifted toward the red, the amount of red shift being on the average proportional to the distance of the galaxy. Initially this was established for relatively bright and therefore near galaxies. Later Hubble showed (1936, 1953) that the red shift relation was applicable to all the galaxies, including the faintest ones, and it was thus recognized as a general law of nature.

The relative shift $\frac{\Delta\lambda}{\lambda}$ of the lines is always constant for a given spectrum, i.e., the wavelengths in each spectrum are changed in the same way as in the Doppler effect. The shift of the entire spectrum can thus be inferred by measuring the shift of a single line, which is highly significant for the weak galaxies where only few spectral lines are discernible.

The most natural and immediate interpretation of the red shift links it up with the Doppler effect. The galaxies are assumed to recede from the observer with velocities which are proportional to the distance:

$$v = Hr. \tag{14}$$

The velocity v entering Eq. (14) should be corrected for the velocity of the Sun relative to the centroid of the nearest galaxies. The latter velocity is mainly determined by the velocity of the Sun in the Galaxy and is equal to about 300 km/sec. It is directed to a point with the coordinates $l = 55°$, $b = 0°$.

The proportionality coefficient in Eq. (14) is called the Hubble constant. Its dimensions are kilometers per second per megaparsec. Hubble's constant is one of the fundamental universal constants and its exact determination is of the utmost significance.

If all the extragalactic objects recede at velocities which are proportional to their distance from the observer, Hubble's law is naturally interpreted as indicating overall expansion of the Metagalaxy: this is the system composed of galaxies and clusters of galaxies, only part of which is accessible to our instruments. Geometrically the red shift law can be interpreted as expansion of space. We must not think, however, that some point near the Sun is the center of the expanding Universe, since in case of isotropic expansion of space the relative recession velocity of any two points is proportional to the distance between them. An inevitable assumption, however, is that the expansion velocity is independent of time, since by observing the distant galaxies we observe their state in the far past.

According to the Doppler law

$$c\,\frac{\Delta\lambda}{\lambda} = v, \tag{15}$$

where c is the velocity of light. Hence, using Eq. (14), we get

$$c\,\frac{\Delta\lambda}{\lambda} = Hr, \tag{16}$$

where again $\Delta\lambda$ should be corrected for the velocity of the Sun relative to the centroid of the nearest galaxies.

Note that Eq. (15) is not exact. According to the special theory of relativity, the Doppler effect is described by the equation

$$\frac{\Delta\lambda}{\lambda}=\sqrt{\frac{1-\frac{v}{c}}{1-\frac{v}{c}}}-1, \tag{17}$$

which reduces to (15) only for small v. The maximum measured red shifts $\frac{\Delta\lambda}{\lambda}$ currently reach 1/3. (Baum (1956) measured with the 200-in. telescope the red shift of a cluster of galaxies corresponding to a velocity of 120,000 km/sec.) In these cases, the relativity corrections in Eq. (15) can no longer be ignored.

Note that although the observations directly lead to Eq. (16), physically it is more logical to start with Eq. (14), and hence using (17) to derive (16) for $\frac{\Delta\lambda}{\lambda}\ll 1$.

Let us now consider the observational proof of Eq. (16). As we know,

$$m-\Delta m-M=5\lg r-5 \tag{18}$$

and using (16) we write

$$m-\Delta m=5\lg c\frac{\Delta\lambda}{\lambda}+(\overline{M}-5-5\lg H)+(M-\overline{M}), \tag{19}$$

where for c in km/sec, H is expressed in km/sec per parsec. The apparent magnitude correction Δm is made up to two parts:

$$\Delta m=\Delta m'+\Delta m'',$$

where $\Delta m'$ is the absorption correction, $\Delta m''$ is the apparent magnitude correction associated with the red shift. The first can be estimated ignoring the metagalactic absorption and evaluating the galactic absorption by any standard technique.

The red-shift correction $\Delta m''$ can be found assuming an average energy distribution in the spectra of all galaxies. Then one calculates the energy distribution corresponding to the observed red shift of all the wavelengths and takes the difference in photographic magnitudes corresponding to the initial and the shifted energy distributions.

Let now H be constant and M independent of the distance (it is thus independent of m). Then by (19) we see that $m-\Delta m$ and $\lg c\frac{\Delta\lambda}{\lambda}$ are related by the statistical equality

$$m-\Delta m=A\lg c\frac{\Delta\lambda}{\lambda}+B, \tag{20}$$

where $A=5$ and $B=\overline{M}-5-5\lg H$, $\overline{M}$ being the mean absolute magnitude of the galaxies.

The most comprehensive check of Eq. (20) was carried out by Humason, Mayall, and Sandage (1956). They used a list of 620 red shifts of galaxies measured at the Mount Wilson and Palomar Observatory and 300 red shifts measured at the Lick Observatory (the Lick list included 114 items figuring in the first list).

The results were applied to plot a diagram of $m-\Delta m$ vs. $\lg c\frac{\Delta\lambda}{\lambda}$ for 474 galaxies which are not members of clusters (Figure 209). The position of the points in this diagram indeed confirms the substantially linear dependence between $m-\Delta m$ and $\lg c\frac{\Delta\lambda}{\lambda}$. We can therefore proceed with a solution of the redundant system of 474 equations (19) with two unknowns A and B. The unknown term $(M-\bar{M})$ in these equations is dropped, since it probably behaves as a random error.

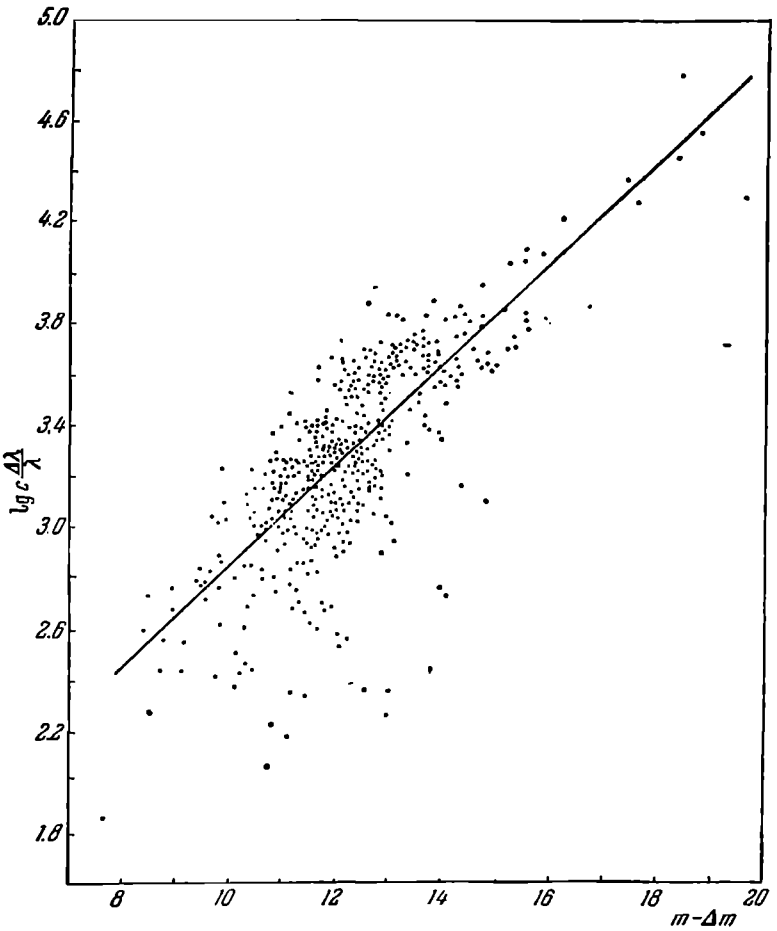


FIGURE 209

The solution by the least squares method gives

$$A = 5.028 \pm 0.116; \qquad B = -4.250 \pm 0.169. \tag{21}$$

A is thus indeed 5, which confirms Hubble's law.

Figure 209 shows the line (20) plotted by the least squares method, i.e., with A and B from (21).

The spread of the points about this line is attributable to the following factors:

1) The absolute magnitudes of the galaxies show a considerable dispersion and the omission of $(M-\bar{M})$ in Eqs. (19) produces a noticeable effect. This is the main source of scatter in the diagram.

2) Alongside with the general law of recession, each galaxy has its own individual velocity. The observed red shift gives the algebraic sum of the space expansion velocity and the individual radial velocity of the galaxy. The individual radial velocities are not known with any accuracy, but the rms velocities are of the order of 200—300 km/sec. The dispersion of the individual radial velocities further enhances the scatter of the points. The distribution of the individual radial velocities is probably symmetric, so that $\lg c\frac{\Delta\lambda}{\lambda}$ should show a skewed deviation from the mean, especially at small distances. This is indeed observed as the large downward deviation of points for small values of $\lg c\frac{\Delta\lambda}{\lambda}$.

3) Many of the near galaxies with measured red shift are dwarfs, whereas no dwarfs are included among the distant galaxies. Thus, for small $c\frac{\Delta\lambda}{\lambda}$ we have an excess of weak galaxies, i.e., galaxies with too high values of $m-\Delta m$. This should also displace the points downward from the straight line in the left part of the plot.

4) Further scatter is caused by errors in the determination of m, $\Delta m'$, $\Delta m''$, and $\lg c\frac{\Delta\lambda}{\lambda}$.

To reduce the spread of points about the red shift line, we should choose galaxies of nearly equal absolute magnitude. This can be achieved, say, by selecting the brightest objects in various clusters. Since these are invariably high-luminosity objects, this choice will enable us to determine the red shift for more distant galaxies, penetrating farther out into space and measuring A and B with higher precision.

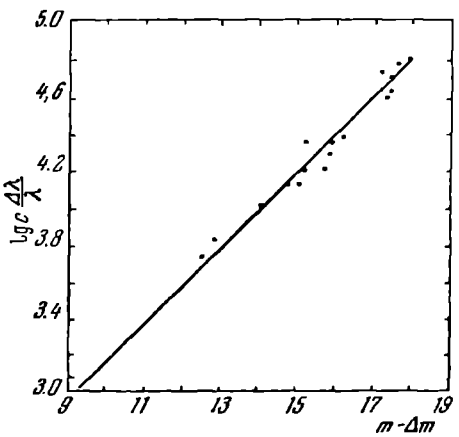


FIGURE 210

Humason, Mayall, and Sandage measured the red shift and the average apparent magnitude for the 1st, 3rd, 5th, and 10th brightest galaxy in each cluster. This technique greatly reduced both the dispersion of the absolute magnitudes and the dispersion of the individual velocities. Eighteen clusters were considered; the minimum radial velocity was 1136 km/sec (the Virgo cluster), and the maximum reached 60,526 km/sec. Figure 210 shows that this selection of material gives a distinct linear Hubble relation and the spread of points about the straight line is insignificant.

The coefficient B obtained by solving the redundant system (20) gives the constant H, if $\overline{M}$ is known. Exact determination of $\overline{M}$ is a difficult problem, however. Therefore H is best determined (once the validity of the Hubble law has been established) directly from Eq. (14) for some particular galaxy, whose radial velocity has been measured and whose distance is known with fair reliability from some alternative method. The various determinations of the constant H are discusssed in §183.

The recession velocities emerging from Hubble's law for the near galaxies are small and the individual velocities may in some cases lead to an apparent negative velocity. Thus NGC 224 approaches our Galaxy with a velocity of 70 km/sec. NGC 598 has an approach velocity of 20 km/sec. These figures have been corrected for the velocity of the Sun relative to the galactic centroid.

§ 183. Revision of the scale of extragalactic distances. Hubble's method for the determination of extragalactic distances (see § 177) gave estimates of the average luminosities of galaxies of various types and made it possible to find the constant H. A whole system, or as we say a scale of extragalactic distances was thus obtained.

Since then, however, the validity of Hubble's scale began to give rise to certain doubts. The main reason for this was that this scale of extragalactic distances led to a little probable conclusion according to which the luminosity and the size of our Galaxy exceeded by a substantial factor the luminosity and the size of all other galaxies. Similarly it was found that the luminosities of certain individual objects in our Galaxy exceeded the luminosities of the corresponding objects in other galaxies. This indicated that the extragalactic distances were definitely too low. Lundmark in 1946 ignored the cepheids and determined the distance to NGC 224 using 110 novae, white supergiants, and 140 globular clusters. His conclusion was that the distance of NGC 224 had to be doubled, since otherwise all the objects included in the sample would be consistently fainter than their analogs in the Galaxy.

To understand the reason behind all this, let us consider in more detail how the scale of extragalactic distances actually developed from the different distance measurement methods.

The most exact method for the determination of the extragalactic distance was the method of cepheids. If the dispersion of the period—luminosity curve of cepheids is small, the accuracy of the method is determined by the accuracy of the zero point.

In other galaxies, only the long-period cepheids are observed. Short-period cepheids (RR Lyrae stars) are invisible because their luminosities are too low. Hence we require the zero point of the period—luminosity curve for the long-period cepheids. Galactic observations, however, could not supply an exact determination of the zero point, since none of the long-period cepheids in the Galaxy are close enought to the Sun to permit measuring the trigonometric parallax. The statistical parallaxes of these stars are determined with considerable uncertainty, again because of their considerable distance from the observer. The zero point of the period—luminosity relation for the long-period cepheids was therefore originally taken to coincide with that of the curve of RR Lyrae stars. The latter was known with considerable accuracy, since some of the RR Lyrae stars are

fairly near. In other words, it was tacitly assumed that the long-period cepheids and the RR Lyrae stars fall along a single curve. Shapley's period—luminosity curve (Figure 211) was widely accepted and generally used in determinations of extragalactic distances by the cepheid method.

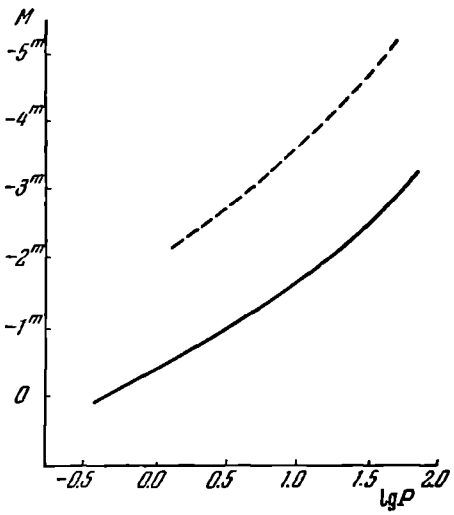


FIGURE 211

Cepheids are visible only in the nearest galaxies. To obtain the distances of farther galaxies, where the cepheids are no longer visible and only the brightest stars are resolved, the apparent magnitudes of these objects in the nearest galaxies were compared and the average absolute magnitudes of the brightest stars were thus determined. Hubble obtained $-6^m.1$ for the average absolute magnitude of the brightest stars. The method of the brightest stars made it possible to fix the position of somewhat more distant galaxies.

To measure the distances to still farther objects, the method of bright stars should be linked up with the method of integrated magnitudes or the method of radial velocities. Since the method of integrated magnitudes is less accurate (because of the considerable disperison of the absolute magnitudes of galaxies), large distances are generally measured by the method of radial velocities. However there is a certain obvious restriction to this technique. The average individual radial velocity of galaxies is approximately 400 km/sec. Therefore, Hubble's constant can reliably be determined only using fairly distant galaxies, whose recession velocity is at least several times greater than the average individual radial velocities, i.e., red shift velocities should be at least 1000 km/sec. In galaxies with $v_r \geqslant 1000$ km/sec, no cepheids are visible. The brightest stars in these galaxies, however, are still visible, although they are accessible to the largest telescopes only. The comparison between the brightest stars and the radial velocities is thus feasible only within a very narrow range of distances, and it can be carried out only for galaxies with radial velocities close to 1000 km/sec. The Virgo cluster was used to this end, and in particular its brightest object, the spiral galaxy NGC 4321 (M100).

The method of brightest stars can be linked up with the method of integrated magnitudes, also using the Virgo cluster. This requires comparison of the apparent magnitude of the brightest stars with the integrated magnitude of the galaxy, assuming that the brightest galaxy in this cluster NGC 4321 (M100) has the same absolute magnitude as NGC 224 and our Galaxy. This comparison was in fact used by Hubble.

Hubble's results gave $H = 540$ km/sec/Mpc.

In summing up we can say that the reliability of the first scale of extragalactic distances was conditioned by the accuracy of three distinct operations: 1) the determination of the zero point of the long-period cepheids, assumed to coincide with the zero point of the RR Lyrae stars; 2) comparison of apparent magnitudes of long-period cepheids and bright stars; 3) comparison of apparent magnitudes of bright stars with radial velocities or integrated magnitudes of galaxies.

The errors in the distances of the nearest objects, where the cepheids are visible, depend only on the zero-point error. Errors in the distances of the farther galaxies, where the cepheids are invisible but the brightest stars are resolved, are determined by the errors under the first two categories. Finally, the error in Hubble's constant, and hence in the distance, as determined from radial velocity is the sum of errors in all the three categories.

In 1954 Baade tried to detect RR Lyrae stars in NGC 224 using the 200-in. telescope. On the old scale, the distance modulus of NGC 224 (not corrected for absorption) was taken to be $m-M = 22^{m}.4$. Since the average absolute magnitude of the RR Lyrae stars is $0^{m}.0$, their apparent magnitude should be near $22^{m}.4$. These extremely faint objects are inaccessible to the 100-in. telescope, whose limiting magnitude is $21^{m}.5$, but they should be visible through the 200-in. telescope, since its limiting magnitude is $23^{m}.0$. However, Baade failed in his attempt to detect any RR Lyrae stars among objects up to $23^{m}.0$. He did obtain the image of long-period Mira Ceti variables ($22^{m}.4$), however. The average absolute magnitude of stars of this type ($-1^{m}.5$) was known with fair certainty from comparison with RR Lyrae stars in globular clusters. Hence, the distance modulus of NGC 224 according to Mira Ceti stars was found to be

$$m - M = 22^{m}.4 - (-1^{m}.5) = 23^{m}.9,$$

i.e., $1^{m}.5$ higher than the previously used distance modulus. If the new distance modulus is correct, the RR Lyrae stars in NGC 224 will have apparent magnitudes of $23^{m}.9$ and this will explain why they did not show in Baade's photographs.

Using the new distance modulus of NGC 224, the absolute magnitudes of novae and globular clusters in this galaxy become equal to the absolute magnitudes of the analogous objects in our Galaxy. This is also a weighty argument in favor of the new value of the distance modulus. However, the adoption of the new distance modulus for NGC 224 should alter the zero point of the period—luminosity curve of the long-period cepheids. If taking $m-M = 23^{m}.9$, we determine the absolute magnitudes of the visible long-period cepheids in NGC 224 and mark them in the diagram of Figure 211, the curve will follow the dashed line.

If we further remember that the right part of the solid curve in Figure 211 was based on the flimsy assumption that the two groups of cepheids followed the same curve, despite their being members of different stellar populations with different physical properties, we finally conclude that the dashed curve in Figure 211 is the correct period—luminosity relation for long-period cepheids, replacing the solid curve which lies $1^{m}.5$ below it.

Since the distances to the nearest galaxies are determined using long-period cepheids, the new period—luminosity relation implies that the

distance moduli of all the nearest galaxies should be increased by $1^m.5$, which is equivalent to doubling the entire scale of distances. The nearest galaxies are thus $1^m.5$ brighter (in absolute terms) and their linear dimensions are double the previous figure.

Baade's adjustment of the distance modulus of the nearest galaxies by $1^m.5$ was confirmed by other astronomers, in particular from observations of variables in Magellanic Clouds.

Note that the distances inside the Galaxy were not affected by this adjustment, since all the large distances in the Galaxy are determined using short-period cepheids, which retained their original period—luminosity relation. The size of the Galaxy, the distance to its center, and the integrated magnitude of the Galaxy therefore remained as before. The only exception was the distance to the rim of the Galaxy, which was usually determined from the farthest long-period cepheids in the direction of the anticenter. This distance was also doubled.

With the revision of the luminosities and the diameters of the nearest galaxies, without changing the luminosity and size of our Galaxy, the Galaxy ceased to be a unique object. It nevertheless remains one of the giant galaxies.

Sandage's recent study (1958) shows that certain additional radical changes should be introduced in the old scale of extragalactic distances. First, proceeding from Haro's result (1956), which points to substantial dispersion of the cepheids about the period—luminosity curve, Sandage determined the distance moduli of the nearest galaxies from novae:

	$m - M$
Large Magellanic Cloud (LMC) / Small Magellanic Cloud (SMC)	19.2
NGC 224	24.6
NGC 598	24.5
NGC 6822	24.1

Comparison of these results with Table 2 shows that the distance modulus increased on the average by $2^m.2$, and not by $1^m.5$ as in Baade's work. This indicates that according to Sandage the old distances of the nearest galaxies should be multiplied by a factor of 2.75, and not 2 as in Baade's results.

Further Sandage noted that the average magnitude of the brightest stars depends on the type of the galaxy. This follow from Table 8, which lists the brightest stars of the nearest galaxies.

TABLE 8

Star	Spectrum	System	M_{pg}
N12 V1 Cyg	B5	The Galaxy	$-9^m.8$
ζ Sco	B1 Ia	"	$-9\ .4$
β Ori	B8 Ia	"	$-8\ .8$
HDE 268970	B2 Ia	LMC	$-9\ .8$
HDE 269781	B9 Ia	"	$-9\ .5$
HD 33579	A2 Ia	"	$-10\ .1$
HD 7583	A0 Ia	SMC	$-8\ .8$
HD 6884	B9 Ia	"	$-8\ .5$
Average absolute magnitude of some of the brightest stars		NGC 224	$-8\ .6$
		NGC 598	$-8\ .9$
		NGC 6822	$-8\ .3$

Indeed, the average magnitude of the three brightest stars in LMC is $0^m.5$ higher than in the Galaxy and $1^m.0$ higher than in the SMC.

Sandage's main contribution was his showing that luminous condensations of H II in the distant galaxies can often be mistaken for bright stars. These H II condensations have a much higher luminosity that the brightest stars, so that the distances of the galaxies may come out too low. Sandage photographed NGC 4321 (M100) on the 200-in. telescope using various filters, and was thus able to differentiate between the brightest stars and H II condensations. The apparent magnitudes of these objects were found to be $20^m.8$ and $19^m.0$, respectively. Hubble had originally used a much weaker telescope without any filters. He therefore mistook the H II condensations for the brightest stars and thus ended up with a distance modulus of the Virgo cluster which was $1^m.8$ too low.

In his comparison of the apparent magnitudes of the brightest stars with radial velocities of the galaxies, Hubble thus introduced a further error of $1^m.8$, in addition to the previous error of $2^m.2$ associated with the zero point of the long-period cepheids. Sandage thus concluded that the distance modulus of the Virgo cluster in the old scale was too low by $4^m.0$. It is therefore not $26^m.7$, as suggested by Hubble, but $30^m.7$. This result was confirmed by Baum (1955), who compared the globular clusters in NGC 224 and NGC 4486 (M87) in the Virgo cluster. Baum found that the distance moduli of the two galaxies differed by $6^m.0$.

If the distance modulus on the old scale is too low by $4^m.0$, Hubble's constant is too high by a factor of 6—7. The true value of H should be 75—90 km/sec/Mpc. In view of the considerable dispersion of the luminosities of the brightest stars in the galaxies (Table 8), Sandage gives a more cautious result, H = 50—100 km/sec/Mpc.

Humason, Morgan, and Mayall (1956) had previous obtained H = = 180 km/sec/Mpc from their data on the red shift of the galaxies, using the revised zero point of the period—luminosity curve for the long-period cepheids. Hubble's first determination of H in the old scale of distances gave 540 km/sec/Mpc.

The revision of the scale of extragalactic distances can be briefly summarized as follow: to obtain the correct distances, the distances of all the galaxies with $m-M$ less than 26^m on the old scale should be multiplied by 2—2.5, and the distance of all galaxies with $m-M$ greater than 26^m on the old scale should be multiplied by 6—7.

To avoid confusion, it is always desirable to indicate the actual scale of distances, wheteher the old scale (Hubble's scale), Baade's scale, or Sandage's revised scale.

§ 184. **Rotation of galaxies.** If a galaxy rotates about an axis which is not parallel to the line of sight, one of the two parts of the apparent ellipse cut by the projection of the spin axis approaches the observer (relative to the center of inertia of the galaxy), while the other part recedes from the observer (again relative to the center of inertia). Thus, different points of the galaxy have different radial velocities in virtue of its rotation.

When the spectrograph slit is projected onto the major radius of the apparent ellipse of a rotating galaxy, it defines a strip with a nonzero gradient of radial velocities, so that the spectral lines will be inclined.

A spectrogram of this kind was first obtained by Slipher in 1914, and the inclination of the spectral lines was correctly interpreted by him as

the result of rotation. The rotation of other galaxies thus had been discovered long before we established the rotation of our star system.

Only the central parts of galaxies have sufficient surface brightness to produce lines on a spectrogram. The inclination of the spectral lines therefore provides an indication of the rotation of the interior regions of galaxies.

Rotation data are currently available for more than 30 spiral and elliptical galaxies. Most of these data were obtained by Mayall.

All the spectrograms show inclined lines, but the lines are invariably straight. Lack of curvature of the spectral lines shows that the galactic nuclei rotate as a rigid body. In rigid-body rotation the radial velocity (relative to the center of inertia) of each point along the major radius of the apparent ellipse is proprotional to the distance from the center of the ellipse, so that in an inclined spectral line the deflection of each point is proportional to the distance from the spectrogram midline.

If the inclination φ of the spectral line is known, we can easily calculate the change in radial velocity per minute of arc along the major radius of the galaxy (given the spectrograph dispersion a in mm per km/sec and the scale of the spectrogram b along the line in mm per min of arc):

$$k = \frac{b}{a} \operatorname{tg} \varphi. \tag{22}$$

Let i be the inclination of the spin axis to the line of sight. Then $\frac{k}{\sin i}$ gives the change in the linear rotation velocity in km/sec per minute of arc, and

$$\frac{k}{\sin i} \frac{3438}{r} \tag{23}$$

is the change in the linear velocity of rotation as we move a distance of 1 pc from the center of the galaxy. Here r is in parsecs, and 3438 is the number of minutes in 1 radian.

To obtain the rotation period of the nucleus, we divide $2\pi \cdot 3.08 \cdot 10^{13}$ by (23):

$$T = 5.63 \cdot 10^{10} \frac{ar \sin i}{b \operatorname{tg} \varphi} \text{ seconds}. \tag{24}$$

According to Eq. (10), the inclination i can be determined if the true sphericity, as well as the apparent sphericity η, are known. For elliptical galaxies the true sphericity is not known and we can only determine the probable value of i, seeing that the true sphericity of E galaxies is at least 0.3 and at most equal to the apparent sphericity. For spiral galaxies, the true sphericity ξ is negligible, and we may take $\sin i = \eta$.

Mayall's spectrograms (1948) show lines with inclinations of up to 5°.4 (probable error 0°.4). The least rotation periods were found to be $2.8 \cdot 10^6$ years for NGC 411 (S0), $6.4 \cdot 10^6$ years for NGC 2683 (Sc), $8.8 \cdot 10^6$ years for NGC 3115 (E7), $10 \cdot 10^6$ years for NGC 3031 (Sb); for NGC 3556 (Sc) and NGC 5055 (Sb) the rotation period of the inner parts is $70 \cdot 10^6$ years and for NGC 7640 (SBc) and NGC 4559 (Sc) the inclination of the spectral lines is so small that the period cannot be measured with any reliability; it is, however, definitely greater than $400 \cdot 10^6$ years. (Mayall's original data were adjusted in accordance with the new scale of distances.) No correlation

was observed between the rotation period of the central parts and the absolute magnitude of the galaxies, although there is a slight dependence between the rotation period and the type of the galaxy. The average rotation period somewhat increases on passing to later types.

For the two nearest spiral galaxies — NGC 224 and NGC 598 — data are available on the rotation of fairly distant peripheral regions. For NGC 224, the measurements were carried out by Babcock (1939), who obtained spectrograms for different parts of the main body of this bright galaxy. NGC 598 (Sc) contains numerous bright condensations whose radial velocities are conveniently measured from their sharp spectral lines. Mayall and Aller (1942) measured the radial velocities of 25 condensations of this type in different parts of NGC 598, including the spiral arms. Since these individual objects also have certain residual velocities, besides their centroid velocity, the measurements were averaged and the results gave the variation of the linear velocity with distance from the center.

Figure 212 which plots the linear rotation velocity of NGC 598 shows that the central part of NGC 598 (about 15') rotates as a rigid body. Outside this core, however, the linear velocity at first increases at a reduced rate and then actually decreases away from the center. The period of rotation of the core of NGC 598 is 120 million years, whereas the peripheral regions rotate with a period of 400 million years.

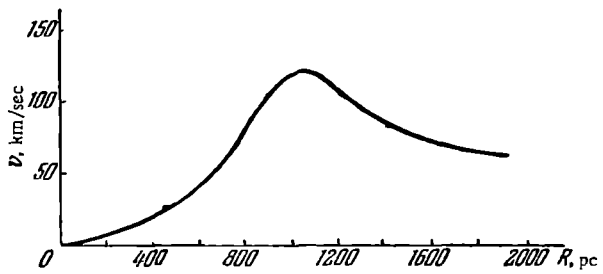


FIGURE 212

The nucleus of NGC 224 rotates with a period of 22 million years, the main central core has a period of 180 million years, and according to Humason a gaseous nebula distant 114' from the center of this galaxy rotates with a period of about 400 million years.

In 1957 van de Hulst, Raimond and van Woerden measured the radial velocities of neutral hydrogen at various points of NGC 224 and obtained the dependence of the rotation velocity on the angular distance from the center (Table 9).

The determination of the sense of rotation of spiral galaxies in relation to their spiral arms is a very difficult problem. Radial velocity observations show which half of the galaxy approaches the observer and which recedes. If i is large and the spiral galaxy is viewed almost edge on (but not exactly edge on), the distribution of dark matter in the galaxy provides and indication which of the two halves is closer to the observer. The spatial

sense of rotation is then determined without difficulty. For large inclinations, however, the spiral arms are not visible. If on the other hand i is small, the spiral arms are excellently visible, but in this case the distribution of dark matter in the galaxy does not show which of the two parts is nearer. The spatial sense of rotation cannot be determined and it thus cannot be compared with the observed configuration of the spiral arms.

TABLE 9

a	Rotation velocity, km/sec	a	Rotation velocity, km/sec
0′	0	60′	263
7′,5	103	75	249
15′	193	90	242
22′,5	241	105	235
30′	267	120	231
37′,5	278	135	226
45′	275	150	221

It is clear, however, that whatever the exact origin of the spiral arms, they should always assume the same orientation relative to the galactic spin. Either the spiral arms in all the galaxies trail (coiling up) or lead (unwinding).

This problem has been repeatedly discussed. The trailing school was represented by Hubble, Holmberg, Vaucouleurs, Irwin, Fricke, Arp, and others. The unwinding hypothesis was advocated by Lindblad (who also proposed a theory of the origin of spiral arms which requires their outward winding) and co-workers and by Baade.

Generally the problem is to determine the configuration of spiral arms in galaxies with an inclination of 60—75° on the one hand and to establish which part is nearer from the distribution of dark matter. It is also necessary to determine which part of the galaxy approaches the observer (from the inclination of the spectral lines).

FIGURE 213

The trailing hypothesis is gradually gaining universal acceptance. As an example, consider the photograph of NGC 7331 (Figure 213), discussed by Vaucouleurs. The distribution of dark matter shows that the lower part of the ellipse is nearer, and the top part is farther. A study of the elements of the spiral arms, especially their tips on the right and the left, shows that the spiral arms in the photograpn are directed counterclockwise. Finally the spectrogram shows that the right part of the galaxy in the

photograph approaches the observer and the left half recedes from the observer (clockwise rotation). Collation of these data indicates that we have here a case of trailing arms.

§ 185. Determination of galactic masses from rotation. Rotation data enable us to estimate another important characteristic, the mass of the galaxies.

First let us consider how to estimate the mass of the nucleus. Observations of galaxies viewed edge on show that the nuclei are not entirely flat: these are objects of fairly pronounced sphericity. Hence it follows that the residual velocities of stars in the nucleus are large and the velocity of an object moving in a circular orbit in the plane of symmetry of the nucleus is substantially greater than the centroid velocity at the same point. Direct determination of the circular velocity for purposes of mass determination from the observed centroid velocity is therefore not an easy task.

The problem should be approached differently. Assuming rigid-body rotation (as the observations indicate), we will consider the galaxies to approximation as homogeneous flattened ellipsoids of revolution in a global quasistationary state. Then, from the theory of rotating quasistationary systems (Agekyan, 1958), we have

$$\varkappa = \frac{25Q^2}{6GM^3R}, \tag{25}$$

where Q is the angular momentum of the nucleus, M is its mass, R is the major radius of the nucleus, and $\varkappa$ is related to the sphericity g of the nucleus by the equalities

$$\varkappa = \sqrt{1+l^2}\left(\frac{3+l^2}{l^3}\operatorname{arctg} l - \frac{3}{l^2}\right); \quad l = \sqrt{g^{-2}-1}. \tag{26}$$

Since for a homogeneous ellipsoid of revolution

$$Q = \frac{2}{5}\omega R^2 M, \tag{27}$$

where ω is the angular velocity of the nucleus, we have from (25) for the mass

$$M = \frac{2R^3\omega^2}{3G\varkappa} \tag{28}$$

and the density

$$\varrho = \frac{\omega^2}{2\pi G\varkappa g} = \frac{2\pi}{G\varkappa g T^2}. \tag{29}$$

The density of the nucleus is thus directly recoverable from the angular velocity and the flattening.

Table 10 lists the values of g and $\varkappa$ for various eccentricities of the meridional sections of the ellipsoid of revolution

Mayall's rotation periods for galactic nuclei lie between $3 \cdot 10^6$ to $70 \cdot 10^6$ years. The observed sphericities of the nuclei range from 0.4 to 0.7. Taking from Table 10 the corresponding values of $\varkappa$, we find from (29) that the density of the galactic nuclei is $10^{-20} - 10^{-22}$ g/cm^3. This is on the average three orders of magnitude higher than the average density of matter near the Sun, i.e., at the periphery of our Galaxy.

TABLE 10

e	g	x	e	g	x
0.0	1.000	0.000	0.96	0.280	0.771
0.4	0.917	0.048	0.98	0.199	0.950
0.6	0.800	0.126	0.99	0.141	1.10
0.8	0.600	0.303	0.995	0.101	1.21
0.9	0.436	0.506	1.000	0.000	$\pi/2$
0.94	0.341	0.656			

A method for the determination of the density in the peripheral regions of galaxies, where differential rotation prevails, was developed by Wyse and Mayall (1945). The galaxy is approximated by an assembly of infinitesimally thin circular disks (the central bulge is ignored when the outer reaches are considered).

Let σ be the surface density of a homogeneous disk, r its radius. From the potential theory we known that the potential of this disk per unit mass at an interior point distant ϱ from the center in the plane of the disk is given by

$$\Phi = 4GrL\sigma, \tag{30}$$

and the potential per unit mass at an exterior point ($\varrho > r$) is

$$\Phi = 4Gr\left[\frac{1}{k}(L-K)+kK\right]\sigma, \tag{31}$$

where

$$L=\int_0^{\frac{\pi}{2}}\frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}} \quad \text{and} \quad K=\int_0^{\frac{\pi}{2}}\frac{\sin^2\varphi\, d\varphi}{\sqrt{1-k^2\sin^2\varphi}}$$

are elliptical integrals of 1st and 2nd kind, whose modulus k is equal to $\frac{\varrho}{r}$ for interior points and $\frac{r}{\varrho}$ for exterior points.

Differentiation of (30) and (31) with respect to ϱ, gives an expression for the attractive force per unit mass:

$$F_c = -\frac{4}{k}G(K-L)\sigma \quad \text{(interior)}, \tag{32}$$

$$F_c = -4G(K-L)\sigma \quad \text{(exterior)}. \tag{33}$$

The minus sign indicates that the attractive force is directed to the center. In what follows this direction will be taken as positive, and the minus sign can be dropped.

Since the density in galaxies is a function of the distance from the center, any point in the plane of the galaxy is attracted by a multiplicity of concentric coplanar disks of equal surface density $d\sigma$ and different radii r.

Since the attractive force due to one disk for an interior point is

$$dF_c = \frac{4G}{k}(K-L)\,d\sigma \tag{34}$$

and for an exterior point

$$dF_c = 4G(K-L)\,d\sigma, \tag{35}$$

the total force due to the multiplicity of infinitesimally thin disks on a unit mass distant ϱ from the center is given by the sum of the integrals of these expressions:

$$F_c = \frac{4G}{k}(K-L)\,\sigma_R + 4G\int_{\sigma_R}^{\sigma_P}\frac{1}{k}(K-L)\,d\sigma + 4G\int_{\sigma_P}^{\sigma_0}(K-L)\,d\sigma, \tag{36}$$

where σ_R is the density at the rim of the galactic disk ($r=R$), σ_P is the density at the particular point, and σ_0 is the density at the center of the galaxy. The first term in Eq. (36) is the attraction of a homogeneous disk of density σ_R and radius R, the second term is the resultant attraction of disks with radii greater than ϱ, and the third term is the resultant attraction of disks with radii smaller than ϱ.

The density σ_R at the rim of the galaxy is zero, and the first term in (36) can be dropped.

The sought quantity is the density. To solve Eq. (36), Wyse and Mayall represented the density as a polynomial in $\frac{r}{R}$:

$$\sigma = \sum_{i=1}^{n} A_i \left(\frac{r}{R}\right)^i, \tag{37}$$

and the problem reduces to the determination of the coefficients of this polynomial.

Seeing that $\sigma_R = 0$, we get

$$\sum_{i=1}^{n} A_i = 0. \tag{38}$$

Further, differentiation of (37) gives

$$d\sigma = \sum_{i=1}^{n} iA_i \left(\frac{r}{R}\right)^{i-1}\frac{dr}{R}. \tag{39}$$

Inserting (39) in (36) we get

$$F_c = 4G\int_R^{\varrho}\frac{1}{k}(K-L)\sum_{i=1}^{n} iA_i\left(\frac{r}{R}\right)^{i-1}\frac{dr}{R} + 4G\int_0^{\varrho}(K-L)\sum_{i=1}^{n} iA_i\left(\frac{r}{R}\right)^{i-1}\frac{dr}{R}.$$

In the first integral $k=\frac{\varrho}{r}$ and in the second $k=\frac{r}{\varrho}$. Changing over to an integration variable k in each of these integrals and putting $\frac{\varrho}{R}=a$, we find

$$F_c = 4G\sum_{i=1}^{n} A_i B_i(a), \tag{40}$$

where

$$B_i(a) = ia^i\left[\int_a^1 k^{-(i+2)}(K-L)\,dk + \int_0^1 k^{i-1}(K-L)\,dk\right]. \tag{41}$$

The functions $B_i(a)$ can be tabulated. Wyse and Mayall tabulated the first five of these functions (Table 11).

TABLE 11

a	$B_1(a)$	$B_2(a)$	$B_3(a)$	$B_4(a)$	$B_5(a)$
0.00	0.000	0.000	0.000	0.000	0.000
0.05	0.152	0.078	0.059	0.053	0.049
0.10	0.250	0.156	0.120	0.106	0.099
0.15	0.327	0.234	0.183	0.161	0.150
0.20	0.391	0.309	0.249	0.219	0.203
0.30	0.488	0.455	0.389	0.344	0.317
0.40	0.556	0.589	0.537	0.486	0.450
0.50	0.600	0.707	0.689	0.646	0.605
0.60	0.622	0.804	0.837	0.819	0.787
0.70	0.622	0.871	0.967	0.994	0.989
0.80	0.598	0.896	1.057	1.143	1.187
0.90	0.544	0.859	1.066	1.211	1.314
0.95	0.497	0.800	1.012	1.172	1.297
1.00	0.416	0.667	0.843	0.978	1.086

Now, writing Eq. (40) for various a, we obtain a redundant system from which the coefficients $A_1, A_2, \ldots, A_5$ can be determined. The sought function σ is then given by (37).

Wyse and Mayall applied their method to determine the variation of density in NGC 224 and NGC 598, for which the circular velocity is known as a function of distance from the center.

In NGC 224, the surface density in the nucleus raches 1500 $M_\odot/\text{pc}^2$, whereas at a distance of 4000 pc and 12,000 pc from the center, the density falls to 400 $M_\odot/\text{pc}^2$ and 10 $M_\odot/\text{pc}^2$, respectively.

To estimate the mass of the entire galaxy, σ should be integrated over the whole disk:

$$M = 2\pi \int_0^R \sigma r\, dr = 2\pi \sum_{i=1}^{n} \frac{A_i}{R^i} \int_0^R r^{i+1}\, dr = 2\pi R^2 \sum_{i=1}^{n} \frac{A_i}{i+2}. \qquad (42)$$

This equation gives $9.5 \cdot 10^{10}\ M_\odot$ for the mass of NGC 224, which is slightly greater than the mass of the Galaxy. The mass of NGC 598 is substantially less, $2.0 \cdot 10^9\ M_\odot$.

If we know the rotation velocity of regions which lie fairly far from the core of the galaxy, we approximately have for these peripheral parts

$$v = \sqrt{\frac{Gm}{r}}, \qquad (43)$$

which corresponds to attraction by a point mass. The inverse conclusion is also true: if starting at a certain distance the linear rotation velocity is inversely proportional to the square root of the distance from the center, the bulk of the galaxy is confined within the corresponding radius and can be simply found from Eq. (43), i.e.,

$$m = \frac{v^2 r}{G}. \qquad (44)$$

The part of the v vs. r curve beyond the point where (43) is satisfied is called the Kepler branch; for NGC 598, for example, the Kepler branch of the curve begins approximately beyond 1400 pc.

Application of Eq. (44) to the bright object at a distance of 13.5 kpc from the center of NGC 224 gave $9.7 \cdot 10^{10} M_\odot$ for the mass of NGC 224, which is in excellent agreement with the result obtained by Mayall and Aller's method.

§ 186. **The Magellanic Clouds and the Andromeda Nebula.** The Large and the Small Magellanic Clouds (LMC and SMC) are our nearest extragalactic neighbors (Figure 214). Their distance from the Galaxy is one third the distance of the next nearest galaxy, that in Sculptor, and one tenth of the distance of the nearest spiral galaxies NGC 224 and NGC 598. This explains the great significance of the Magellanic Clouds, since their proximity makes it possible to identify and study individual high-luminosity stars and other objects through moderate telescopes. Weak objects, which are so far inaccessible to observations in other star systems, are resolved by the large telescopes in the Magellanic Clouds. Certain objects are actually easier to observe in the Magellanic Clouds than in our Galaxy, since comparison of physical characteristics of different specimens is facilitated by the constancy (not perfect, of course) of their distances and the constancy (again not perfect) of absorption. The Magellanic Clouds, however, are exceptionally interesting galaxies in their own right, vecause of the great variety and richness of their stellar population.

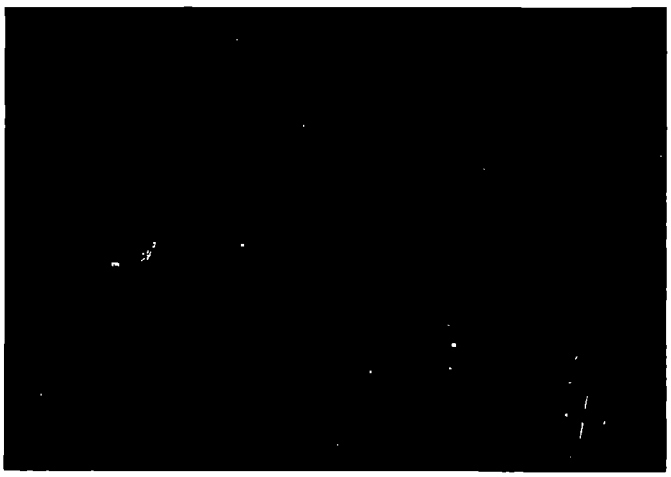


FIGURE 214

The first European to describe the Magellanic Clouds was Magellan's companion, Pifagetta. John Herschel observed the Clouds in the 1830's in South Africa. In 1867 Abbé first suggested that the Magellanic Clouds were two near galaxies. In 1910 Miss Leavitt, while studying the variable stars in the LMC, discovered the period—luminosity relation of cepheids, which since then has become an indispensable tool of galactic and extragalactic research.

The position of the Magellanic Clouds in the Southern Hemisphere interfered with the detailed study of these galaxies, since there are but few observatories in the South. Recently the new Southern observatories have devoted considerable efforts to rectify the earlier omission.

According to John Herschel, LMC measures $7° \times 6°$, and SMC $3°.5 \times 2°.5$. The separation between their centers is 21°. Long-exposure photographs taken with high-speed instruments and star counts give substantially larger dimensions.

Vaucouleurs (1955) counted the stars up to $14^m.0$ in LMC. These are supergiants with absolute magnitudes of $-4^m.7$ and brighter. They occupy a region measuring $12° \times 12°$, and half of the 4700 stars of this group are concentrated within a circle of radius 5°.

In the Small Magellanic Cloud, the stars up to $14^m.3$ and up to $16^m.0$ were counted. There were 500 stars up to $14^m.3$ ($M < -4^m.4$) within an area of $9° \times 4°$, half of which fell inside a circle of radius 1°.2. Stars of up to $16^m.0$ ($M < -2^m.7$) numbered about 10,000.

The Magellanic Clouds are the only extragalactic nebulae where the short-period RR Lyrae stars can actually be compared with long-period cepheids, novae, globular clusters, and other objects used in distance determination. The Magellanic Clouds thus actually fix the scale of extragalactic distances.

Table 12 lists the distance moduli of the Magellanic Clouds obtained by Vaucouleurs for various groups of constituent objects.

TABLE 12

Method	LMC	SMC
Lo. g-period cepheids	$19^m.05$	$18^m.85$
Long-period cepheids with $P \sim 5$ days	18 .65	18 .85
RR Lyrae stars	18 .9	18 .8
Bright stars in globular clusters	18 .7	18 .8
Integrated magnitudes of globular clusters	18 .2	18 .2
Novae at the maximum	19 .0	18 .6
Novae 15 days after the maximum	18 .4	18 .8
Weighted average	18 .72 ± 0.12	18 .75 ± 0.09

We see that the distances obtained for various groups of objects show a good fit, and the mean error of the distance modulus is not large. The distance moduli should be corrected for absorption of light in the Magellanic Clouds and in the Galaxy. The absorption correction is estimated at $0^m.45$, and for the two Magellanic Clouds we thus have

$$m_0 - M = 18.3 \pm 0.1, \tag{45}$$

where the error of the absorption correction has not been included in the mean error.

The distance modulus (45) corresponds to the new scale of extragalactic distances. It gives a distance of 46 kpc, i.e., approximately twice the distance given in Table 1, where the old scale was used.

Vaucouleurs obtained a new estimate of the integrated brightness of the Magellanic Clouds, allowing for the peripheral regions as well. He measured the surface brightness of 621 areas covering a total of 288 sq. degrees in LMC and 252 areas covering 53 sq. degrees in SMC. The isophotes were plotted to bring out the shape of the galaxies and the integrated apparent magnitude was determined. The absolute magnitude was then easily calculated using the distance modulus $18^m.3$ (Table 13).

TABLE 13

Magellanic Clouds	m_b	m_v	$m_b - m_v$	M_b	M_v	Luminous area
Small	+0.75	0.30	+0.45	−17.55	−18.00	200 sq. deg^2 = = 128 kpc^2
Large	+2.70	+2.40	+0.30	−15.60	−15.90	25 sq. deg^2 = =16 kpc^2

Thus the apparent magnitude of the Magellanic Clouds listed in Tables 1 and 3 is substantially higher than the true magnitude, since the tables only give the integrated apparent magnitude of the central part of these galaxies. Particularly pronounced are the differences in the absolute stellar magnitudes of the Magellanic Clouds in Table 13 and 3, since the apparent magnitude correction and the scale correction reinforce one another.

Vaucouleurs gives 20° × 20° for the angular size of the LMC and 9° × 4° for the SMC. This corresponds to linear dimensions of 15 × 15 kpc and 6 × 3 kpc.

The maximum surface brightness in the central parts of the LMC was found to be 20.53 stellar magnitudes per sq. second, and in the SMC 21.10 stellar magnitudes per sq. second.

From an examination of the isophotes of the Magellanic Clouds and the distribution of stars in these galaxies, Vaucouleurs identified traces of spiral structure. He also concluded that the two galaxies were relatively highly flattened. This is clearly an indication of their rotation. The equatorial plane of the LMC makes an angle of 65° with the line of sight, and that of SMC 30°. The Small Magellanic Cloud is thus viewed almost edge on.

The galaxies NGC 55, NGC 4618, and NGC 4625 are similar to the Magellanic Clouds in their appearance. Galaxies of this type, classified as I II or (in Vaucouleurs's classification) Sm, occur either in pairs or jointly with a galaxy of some other type, generally a spiral.

The distance of the Magellanic Clouds from the Galaxy is only three times the diameter of the LMC and less than double the diameter of the Galaxy. There is thus a definite possibility of interaction between our star system and the Magellanic Clouds. John Herschel was the first to point out that the Milky Way and the LMC are joined by a lane of matter. Vaucouleurs noted that the isophotes of the Milky Way constructed by Pannekoek and Koelbloed show a kind of a luminous prominence in the direction of LMC at galactic longitude 230°. LMC in its turn has a sort of a spiral branch which extends away from the Galaxy, so that it should also have a similar branch extending toward the Galaxy. The Large Magellanic Cloud and the Galaxy are thus probably linked by a bar.

In terms of stellar composition, LMC and SMC are galaxies with prevalence of Population I stars. SMC also has a substantial number of Population II stars, whereas their proportion in LMC is negligible. This difference is consistent with the much higher content of dust matter in LMC than in SMC. Shapley (1942) showed that the number of distant galaxies which show through the LMC is less than the usual number of galaxies, reaching the normal average only at a distance of $4-6°$ from the geometrical center of the Cloud. Hence the conclusion that LMC contains dark absorbing matter in large quantities. The number of distant galaxies showing through SMC is equal to the normal average.

The Large Magellanic Cloud contains an abundance of remarkable specimens of Population I stars, and in this respect is it clearly superior even to the spiral arms of our Galaxy. In LMC there are great numbers of blue supergiants, including three stars with absolute magnitude less than -9^m and one star (S Dor) with absolute magnitude less than -10^m. There are also red supergiants. Five of these have visual absolute magnitudes close to -9^m and spectral types K0—M0 (Feast and Thackeray, 1956). According to these authors, the list of the brightest stars in the Magellanic Clouds apparently includes all the blue stars, but it is far from being complete in regard to red stars, since it is extremely difficult to establish what red stars actually belong to the Magellanic Clouds. Among the large gaseous nebulae we should mention 30 Dor, which greatly exceeds in size and luminosity the gaseous nebulae in the Galaxy. Its absolute magnitude is approximately -15^m, i.e., its luminosity is greater than that of some galaxies; its diameter is about 200 pc.

Kerr and Hindman (1953) used a 9-m radio telescope to study the neutral hydrogen emission at 21 cm from the Magellanic Clouds. They determined the intensity and the average wavelength at more than 200 points. The average wavelength gives the radial velocity. The radio isophotes of LMC turned to be close to the optical isophotes, but in SMC the radio isophotes revealed an extensive hydrogen atmosphere, of much higher intensity and extent that anything suggested by the optical observations. The analysis of radio isophotes gave $6 \cdot 10^8 M_\odot$ for the mass of hydrogen in LMC and $4 \cdot 10^8 M_\odot$ for that in SMC. The density of hydrogen in the central regions of LMC and SMC is about 1 atom/cm^3, which is approximately equal to the density of hydrogen in the spiral arms of the Galaxy.

The radio wavelengths give + 280 km/sec for the radial velocity of LMC as a whole and 160 km/sec for that of SMC as a whole. These velocities show a good fit with the radial velocities obtained by optical methods for the hydrogen nebulae. This fit primarily establishes the accuracy of the measurements and also confirms the applicability of the Doppler law to all wavelengths of the electromagnetic spectrum. The measured velocity incorporates the velocity of the Sun in the Galaxy. Radio estimates give 220 km/sec for the latter, and it points almost directly away from the Magellanic Clouds. Correcting for this velocity, we obtain + 82 km/sec for the velocity of LMC relative to the Galaxy. The corresponding velocity of SMC is + 18 km/sec.

Measurements of the neutral hydrogen wavelengths at different parts of the Magellanic Clouds enabled Kerr and Vaucouleurs (1956) to determine the rotation of LMC. The curve of linear velocity vs. distance from the center is reminiscent of the analogous curves for other galaxies. The mass

of LMC estimated from rotation data is $(4-5)\cdot 10^9 M_\odot$. The mass of SMC is more difficult to estimate. Kerr and Vaucouleurs suggest that it is one third of the LMC mass.

The radio emission of the ionized hydrogen from Magellanic Clouds at 3.5 m wavelength was studied by Mills. The intensity of this radiation is small over most of the area occupied by these star systems, and it cannot be separated from the 3.5 m emission of the Galaxy. In those regions where the two sources of this radiation can be separated, no correlation is found between the radio intensity and the optical surface brightness. The only exception is the area of 30 Dor, where a sharp radio peak is observed.

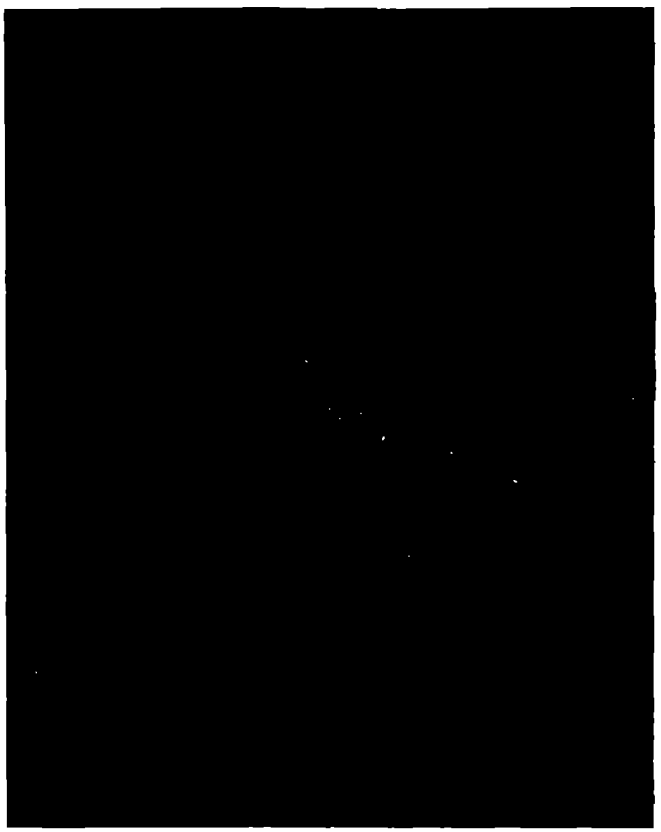


FIGURE 215

The spiral galaxy in Andromeda NGC 224 (M31) is the brightest galaxy in the Northern Sky (Figure 215). It is described as Hubble's type Sb and Morgan's type kS5. Its distance modulus is $23^m.9$ (the new scale), and correction for light absorption gives $23^m.3$. This corresponds to a distance of 460 kpc. NGC 224 has a large bright nucleus with marked central concentration. The nucleus dimensions according to Danver (1942) are

9'× 7', which corresponds to linear dimensions of 1200× 900 pc. The angular dimensions of the entire galaxy, measured by Hubble on a photograph, are 160'× 40', i.e., 20× 5 kpc. In 1934 Stebbins and Whitford determined the limits of NGC 224 photoelectrically and obtained 450'× 110' for the angular dimensions, so that the linear dimensions are substantially greater than those of the Galaxy (60× 15 kpc). It seems, however, that the high-sensitivity photoelectric method fixes the outermost reaches of NGC 224, the equivalent of which in the Galaxy cannot be registered by measuring the distances of high-luminosity objects in the direction of the anticenter. The actual size of NGC 224 is apparently not much greater than the size of the Galaxy. This conclusion also follows from a comparison of the integrated magnitudes. The absolute magnitude of NGC 224 is $-19^m.6$. The Galaxy approximately has the same absolute magnitude. Note, however, that the integrated apparent magnitude of NGC 224 was determined over a smaller region than that outlined by Stebbins and Whitford, so that the resulting luminosity is too low. On the whole it seems that NGC 224 somewhat exceeds the Galaxy both in terms of luminosity and size, while both these systems are classified as giants.

The spin axis of NGC 224 makes an angle of some 75° to the line of sight, so that it is viewed almost edge on. Nevertheless, the 75° angle is not large enough for the dark matter lying in the principal plane of the galaxy to show as a black dust lane. The dust collected into isolated clouds and groups of clouds shows in the form of dark patches in the spiral arms. The spiral arms are moderately developed, they closely hug the nucleus and have certain features which appear to point to traces of ring structure. This impression, however, may be due to the edge-on attitude of NGC 224.

According to the general rule for spiral galaxies, the nucleus of NGC 224 is redder than the galaxy as a whole, and the spiral arms are bluer. The brightest objects in the nucleus are the red supergiants with absolute magnitude of $-3^m.0$. The Hetzsprung—Russell diagram of the nucleus is reminiscent of the diagram of globular clusters. In the spiral arms the brightest stars are the blue supergiants with absolute magnitudes around -8^m. However, in terms of relative abundance of the blue supergiants, NGC 224 is inferior both to LMC and the Galaxy.

In 1955 Baade and Sweeny studied with the 100-in. telescope the distribution of variable stars in NGC 224 in three areas, distant 15' (I), 35' (II), and 50' (III) from the nucleus. The count of variable stars up to $22^m.8$ in these areas gave 150, 205, and 360, respectively. Table 14 lists the frequency of different types of variables in two of these areas.

TABLE 14

Variables	I	%	III	%
Long-period cepheids	38	32	237	70
Eclipsing variables	5	4	30	9
Irregular variables	32	27	20	6
Long-period variables	8	7	2	1
Novae	7	6	1	0
Unidentified	30	24	50	14

We see that novae explode mainly in the part of the spiral arms adjoining the nucleus, where the number of irregular variables is also relatively high. Long-period cepheids, on the other hand, are concentrated mainly in the outer region of the spiral arms.

Arp (1956) carried out a detailed study of novae in NGC 224. During 290 nights spread over a year and a half he took some 1000 photographs through the 60-in. reflector covering an area of $1° \times 3°$ along the major axis of NGC 224. He detected 30 novae with apparent maximum magnitudes between $15^m.8$ and $18^m.2$. The yearly frequency of novae in the entire galaxy was correspondingly estimates as 26 ± 4.

Using the 200-in. telescope and 60—80 min exposures, Baade photographed five planetary nebulae in NGC 224 with apparent magnitudes from $21^m.7$ to $22^m.2$.

TABLE 15

a	$100\,\bar{N}$	a	$100\,\bar{N}$
0′	38	90′	5
15	5	105	7.5
30	7	120	3.5
45	7.5	135	0
60	28.5	150	1.5
75	3		

NGC 224 contains 140 globular clusters. In the new scale of extragalactic distances, their absolute magnitudes and diameters are comparable with those of the globular clusters in the Galaxy.

Hiltner (1958) investigated the polarization of light from 21 of these globular clusters. The four globular clusters with the most reliable polarization measurements show that the electric field vector in NGC 224, like that in the Galaxy, is roughly parallel to the principal plane of the star system.

NGC 224 is surrounded by four elliptical galaxies, two near (NGC 221 and NGC 205), and the other two, NGC 147 and NGC 185, somewhat more distant. The luminosities of these galaxies are substantially less than the luminosity of NGC 224. Their mass is apparently also smaller; these elliptical galaxies may therefore be treated as satellites of NGC 224.

Van de Hulst, Raimond, and van Woerden (1957) investigated the distribution of the 21-cm radiation from NGC 224 and determined the density of neutral hydrogen as a function of the angular distance from the center (Table 15). In the table, the dimensions of $\bar{N}$ are $cm^{-3} \cdot kpc$, so that in order to obtain the surface density per cm^2 $\bar{N}$ should be multiplied by $3.08 \cdot 10^{21}$.

The distribution of hydrogen in NGC 224 is not unlike that in the Galaxy. The total mass of neutral hydrogen in NGC 224 is about $2.5 \cdot 10^9$ solar masses.

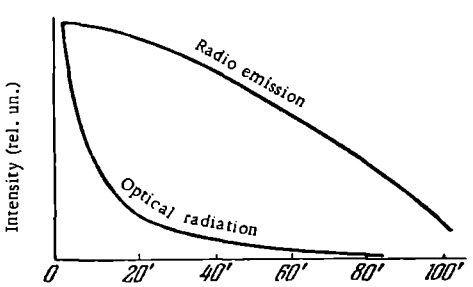


FIGURE 216

Baldwin (1954) measured the intensity of 3.7-m radio emission of NGC 224 along various radii of the apparent ellipse. The resulting curve of relative intensities in optical and radio waves along a radius very close to the minor axis of the apparent ellipse (Figure 216) shows that the radio intensity falls off at a much slower rate than the optical radiation does. The radio image of NGC 224 in this direction thus extends much farther out. It thus follows that the radio sources constitute a spherical subsystem. A similar result was obtained by Brown and Hazard (1951). In 1952 Shklovskii studied the distribution of radio intensity in our Galaxy and established that it was surrounded by a quasispherical radio corona. This result for our Galaxy is consistent with the observations of NGC 224.

§ 187. **Apparent distribution of galaxies.** In 1932 Shapley and Ames published a catalogue of all galaxies up to $13^m.4$ over the entire sky. It listed 1249 galaxies, and comprehensiveness was guaranteed up to $12^m.6$. In addition to photographic magnitude and equatorial and galactic coordinates, the catalogue gives the angular diameter and the type of each galaxy.

At about the same time Shapley and co-workers launched an enormous project involving counts of galaxies up to 18^m in areas covering the entire sky. This project has not been completed yet, but parts have already been published. For instance, counts of galaxies according to magnitudes in 2175 areas covering the part of the sky with $\delta < -60°$ were published in 1937.

In 1934 Hubble published counts of galaxies up to 20^m in 1283 areas from photographs taken through the 60-in. and 100-in. Mount Wilson reflectors. Of these areas, 760 were uniformly distributed over the part of the sky with $\delta > -30°$, and the remaining areas were selected in various critical locations (also in the region $\delta > -30°$). Hubble's counts show that the number of galaxies $N(m)$ up to a given magnitude m in 1 sq. degree of the sky is described by the equality

$$\lg N(m) = 0.6m - 9.12. \qquad (46)$$

In the derivation of Eq. (46), Hubble introduced a correction for the red shift in the galactic spectra. This correction is absolutely essential, since the red shift displaces the intensity peak of the galaxies from the yellow region (average spectral type G) to the red region, where the photographic plates are little sensitive. The photographic magnitudes of rapidly receding, i.e., distant and weak galaxies are additionally increased.

The coefficient 0.6 before m shows (see § 116) that up to distances corresponding to the mean distance of galaxies of apparent magnitude 20, the galaxies on the average are uniformly distributed in space and light absorption is insignificant.

Hubble's result was considered for a long time as an indication of the uniform distribution of galaxies throughout the Universe. However, as we have seen in § 116, Seeliger's inverse theorem is not true. Galaxies may collect into groups and clusters, and these may combine into higher-order clusters, and Eq. (46) will nevertheless remain valid on the average. As we shall see, this is indeed the case. The insignificant absorption of light over distances corresponding to the average distance of the 20-mag galaxies, on the other hand, is a fairly certain fact.

Another important conclusion of Hubble was that the number of galaxies per unit sky area on the average regularly decreases toward the galactic

equator, and virtually no galaxies are visible at the galactic equator. This observed effect, the opposite of galactic concentration, can hardly be regarded as a significant reflection of the true distribution of galaxies in space. It should rather be attributed to the absorption of light in the Galaxy. Indeed, if we take the density of dust matter to be only a function of the distance from the galactic plane and assume the Sun to lie in this plane, we obtain for the absorption of light from extragalactic objects by Eq. (52) of Chapter XVIII

$$\Delta m = \int_0^\infty a(z)\,dr = \int_0^\infty a(z)\,\operatorname{cosec} b\,dz = \operatorname{cosec} b \cdot \beta, \tag{47}$$

where

$$\beta = \int_0^\infty a(z)\,dz \tag{48}$$

is the absorption, in stellar magnitudes, in the direction of the galactic pole.

Now, according to Eq. (15) of Chapter XVIII, assuming the conditions of Seeliger's theorem to hold true, we have

$$N(m-\Delta m) \approx N(m)\cdot 10^{-0.6\Delta m} \tag{49}$$

Since in the presence of absorption Δm objects of magnitude $(m-\Delta m)$ are registered as objects of magnitude m, and the absorption is a function of b, we may write

$$N(m-\Delta m) = N_b(m).$$

Using (47) and extracting the logarithm, we get

$$\lg N_b(m) = \lg N(m) - 0.6\cdot\beta\cdot\operatorname{cosec} b. \tag{50}$$

Here $N(m)$ is the number of galaxies up to m-th magnitude which would have been observed in the absence of light absorption.

Hubble found that his counts of galaxies showed a good fit with Eq. (50) where $\beta = 0^m.25$ and $\lg N(20) = 2.115$, without correction for the red shift.

Note that Hubble's figure $0^m.25$ for the absorption in the direction of the galactic poles is much too low. At the time of Hubble's work the absorbing medium was regarded as continuous and essentially homogeneous, with occasional isolated "coal sacks" — massive dark clouds — dispersed throughout. Hubble's aim was to check the absorption of the homogeneous medium, and he therefore selected areas presumably without any coal sacks. All areas with galaxy counts below the average were thus automatically disregarded. They accounted for 168 of the total number of 760 uniformly distributed sky areas.

In the light of modern notions of the structure of interstellar medium, we realize that the 168 areas discredited by Hubble are directions characterized by positive fluctuations in the number of dark clouds intercepting the line of sight. Clearly, by omitting these areas from his analysis, Hubble obtained a low value of β. If all Hubble's areas are used, we get $\beta = 0^m.458 \pm 0^m.015$, $\lg N(20) = 2.236 \pm 0.021$. The increase in $N(20)$ is also

understandable: the absorption is higher than Hubble's figure, so that correcting for this absorption we end up with more numerous observed galaxies. The determination of β by a number of other authors using different observational material gave intermediate values:

$0^m.305$ (Williams, 1934)
$0^m.31$ (Oort, 1938)
$0^m.34$ (Parenago, 1940)
$0^m.23 - 0^m.40$ (Markaryan, 1946)
$0^m.34$ (Vashakidze, 1953)

Currently, Shane and Wirtanen are engaged on a large-scale project which calls for galaxy counts up to 18^m in the entire part of the sky with $\delta > -23°$ in areas of 1 sq. deg. Counts in individual areas covering millions of galaxies were already published for $[12^h \leqslant \alpha \leqslant 18^h;\ -20° \leqslant \delta \leqslant +20°]$ (1954) and $[0^h \leqslant \alpha \leqslant 6^h;\ +20° \leqslant \delta \leqslant +60°]$ (1956). Herzog, Wild, and Zwicky are in the process of compiling a catalogue of coordinates and apparent magnitudes of galaxies up to $15^m.5$ north of $\delta = -30°$. This catalogue will eventually comprise some 35,000 galaxies.

Comparison of the counts of galaxies in adjoining sky areas which lie at the same galactic latitude shows fluctuations between considerable limits, which greatly exceed the level of natural fluctuations. Two hypotheses have been advanced to account for this fact: 1) the observed fluctuations in the distribution of galaxies are associated with random fluctuations in the distribution of dark clouds in the Galaxy, and 2) the observed fluctuations are attributed to the tendency of galaxies to collect in clusters.

The first hypothesis was investigated by Ambartsumyan over the period from 1940 to 1951, when he laid down the foundation for the theory of fluctuations in statistical astronomy. The second hypothesis was developed by Neyman and Scott (1952), Shane (1953), and others.

The dark matter in the Galaxy is now known to be distributed in patches. It is also known, on the other hand, that the galaxies do form clusters. Both factors thus operate, and the two hypotheses must be considered jointly. Ambartsumyan's generalized method is suitable for this purpose.

We will use Eq. (87) derived in § 129. First we apply this equation to an observer which studies the surface brightness of the Metagalaxy from the rim of our Galaxy. Suppose that all the galaxies are collected into clusters, i.e., there is no continuous distribution of luminous matter, so that $\lambda \neq 0$ and $\eta = 0$. It is further assumed that the distribution function of the intensity of bright clusters, i.e., the normalized distribution function of galaxies in clusters, is the same at any point in space. The presence of dark matter in the Metagalaxy is ignored, i.e., we take $\tau = 0$, $\varkappa = 0$. Eq. (87) then takes the form

$$\frac{d\mu_n^*(s)}{ds} + \lambda(s) \sum_{k=1}^{n} C_n^k \bar{\xi}^k \mu_{n-k}^*(s) = 0, \tag{51}$$

where the moments are superscripted with asterisks, since these are not moments that we can obtain from observations (our hypothetical observer is situated outside the Galaxy, after all). Examination of Eq. (51)

shows that the moments entering this equation are bounded functions only if $\lambda(s)$ goes to zero for $s \to \infty$ faster than $\frac{1}{s^{1+\sigma}}$, where σ is an arbitrarily small positive number. This is associated with the so-called Olbers paradox, which is described in § 193.

Since in the following we will have to change over from fluctuations in the surface brightness of the Metagalaxy to fluctuations in the number of galaxies up to a given magnitude, we may take $\lambda = 0$ for all $s > R$, since all the galaxies up to a given magnitude are confined inside a sphere of some radius R. This ensures a finite μ_n^* (and resolves the Olbers paradox). Treating Eq. (51) as a linear differential equations in $\mu_n^*(s)$, we obtain its solution in the form

$$\mu_n^*(s) = \sum_{k=1}^{n} C_n^k \bar{\xi}^k \int_s^R \lambda(s)\, \mu_{n-k}^*(s)\, ds, \tag{52}$$

where the integration constant is determined from the condition $\mu_n^*(s) \to 0$ for $s \to R$: as long as galaxies lying outside a sphere of radius R are not counted, they are nonexistent to all purpose and intent, so that the "observed" surface brightness outward from the boundary of this sphere is zero.

If we further assume that all the galaxies have the same luminosity, and the average distribution density of their clusters in a sphere of radius R is constant at all points with $s \ll R$, we obtain $\lambda = \text{const}$ and Eqs. (52) for $s = 0$ (i.e., for an observer on the rim of the Galaxy) take the form

$$\left.\begin{aligned} &\mu_1^*(0) = \lambda R \bar{\xi}, \\ &\mu_2^*(0) = \lambda^2 R^2 (\bar{\xi})^2 + \lambda R \bar{\xi^2}, \\ &\mu_3^*(0) = \lambda^3 R^3 (\bar{\xi})^3 + 3\lambda^2 R^2 \bar{\xi}\,\bar{\xi^2} + \lambda R \bar{\xi^3}. \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}\right\} \tag{53}$$

Let us now consider the brightness fluctuations of the Metagalaxy as observed from a point inside the Galaxy. Since we are concerned with the brightness of the Metagalaxy (later on we will change over to the number of observed galaxies per unit sky area, i.e., a quantity which can in no way be confused with observed objects of galactic origin), the line of sight crossing the Galaxy does not intercept any luminous matter of the relevant kind. In Eq. (87) from § 129 we should now take $\eta \equiv 0$, $\lambda \equiv 0$. Now, in accordance with the data of Chapter XVIII we assume that all the dark matter in the Galaxy is collected into discrete clouds, and there is no continuous substrate. Then $\varkappa \neq 0$ $\tau \equiv 0$. Eq. (51) takes the form

$$\frac{d\mu_n(s)}{ds} - \varkappa(s)(1 - \bar{q^n}) \cdot \mu_n(s) = 0. \tag{54}$$

Solving this homogeneous linear differential equation, we should remember that as s increases and we move outside the Galaxy, our problem develops to the previously considered case of fluctuations in the brightness of the Metagalaxy as viewed by an observer outside the Galaxy. Therefore, remembering the tremendous disparity in the scales of the Galaxy and the Metagalaxy, we may write

$$\mu_n(\infty) = \mu_n^*(0). \tag{55}$$

Using (55) to determine the integration constant in the solution of (54), we get

$$\mu_n(s) = \mu_n^*(0)\, e^{-(1-\overline{q^n})\int_s^\infty \varkappa(s)ds} \tag{56}$$

If $\varkappa(s)$ is only a function of the distance from the plane of the Galaxy, we find as in (47)

$$\int_0^\infty \varkappa(s)\, ds = \xi \operatorname{cosec} b, \tag{57}$$

where b is the galactic latitude, and ξ is the expectation value of the number of dark clouds in the direction to the galactic pole. Then from (56) for $s = 0$ (i.e., an Earth-bound observer) we get

$$\mu_n(0) = \mu_n^*(0)\, e^{-(1-\overline{q^n})\xi \operatorname{cosec} b}. \tag{58}$$

It now remains to change over from the random variable "surface brightness" to another random variable "surface density of the number of galaxies." We recall that according to Pogson's rule an increase of Δm in apparent magnitude involves a reduction in brightness by a factor of $10^{0.4\cdot\Delta m}$. Comparison with (49) thus shows that, assuming screening by dark matter, the surface brightness x and the number of galaxies N are related by

$$N \sim x^{\frac{3}{2}}.$$

At the same time, for the fluctuations associated with clusters of galaxies we have

$$N \sim x.$$

Therefore, on changing over from Eq. (58) to the corresponding equations for the moments of the surface density of galaxies V_n, the power exponent n of q should be replaced with $3n/2$. Extracting the logarithm and inserting for μ_n^* (the surface density moments are designated V_n^*) their expressions from (53), we thus obtain for $n = 1, 2, 3, \ldots$

$$\left.\begin{aligned}
&\lg V_1(b) = \lg(a\bar{g}) - \alpha(1-\overline{q^{1,5}})\,\xi \operatorname{cosec} b,\\
&\lg V_2(b) = \lg[a^2(\bar{g})^2 + a\overline{g^2}] - \alpha(1-\overline{q^3})\,\xi \operatorname{cosec} b,\\
&\lg V_3(b) = \lg[a^3(\bar{g})^3 + 3a^2\bar{g}\overline{g^2} + a\overline{g^3}] - \alpha(1-\overline{q^{4,5}})\,\xi \operatorname{cosec} b,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}\right\} \tag{59}$$

where α is the conversion factor from Napierian logarithms to decimal logarithms, $a = \lambda R$ is the expectation value of the number of clusters within a sphere of the radius R, and $\bar{g}$, $\overline{g^2}$, etc., are respectively the mean number of galaxies in a cluster, the mean square of the number of galaxies in a clusters, and so on.

The left sides of Eqs. (59) are obtained from observations. This is the mean value of the number of galaxies at latitude b raised to first, second,

etc., power. Since these observations can be made for different b, each equation in (59) gives a redundant system of equations. Solving these redundant systems, we obtain $a\bar{g}$, $(1-q^{1.5})\xi$; $a^2(\bar{g})^2+a\bar{g}^2$, $(1-\bar{q}^3)\xi$, etc.

Note that this method enables us to distinguish between the clustering effect, which is represented by the first terms on the right in Eqs. (59), and the effect of patchy absorption, accounted for by the second terms on the right.

Applying the first two equations in (59) to Shapley's calculations (1937), Agekyan (1957) obtained the following results: the average number of clusters of galaxies up to $17^m.2$ per 1 sq. deg is $a = 2.18 \pm 0.02$; the average number of galaxies in a cluster is $\bar{g} = 8.3 \pm 1.6$; average absorption of one dark cloud $\Delta m = -2.51 \lg \bar{q} = 0^m.39 \pm 0^m.02$.

The clusters are thus mainly small aggregates containing on the average 8 non-dwarf galaxies (only the nearest dwarf galaxies are visible, and they virtually do not enter Shapley's calculations).

In 1953 Neyman, Scott, and Shane applied their theory of fluctuations in the surface density of galaxies to the results of Shane and Wirtanen (1954) and assuming the fluctuations to be caused only by the clustering effect they obtained $\bar{g} < 287$. Comparison with the previous value of $\bar{g}$ shows that the effect of patchy absorption on the apparent distribution of galaxies is most significant.

In conclusion note that according to Zwicky the fluctuations in the apparent distribution of galaxies are due to some other factors besides the clustering effect and the patchy absorption. In his opinion, the tendency of clusters of galaxies to aggregate into higher-order clusters is of the utmost significance. Fluctuations in metagalactic absorption also make a substantial contribution. These factors may indeed prove quite significant for exceedingly weak distant galaxies.

§ 188. Double and multiple galaxies. Galaxies reveal a pronounced tendency to form double and multiple systems. The nearest pair of galaxies — the Large and the Small Magellanic Clouds — constitute a double system, and together with our Galaxy they aggregate into a triple system. The Andromeda Nebula with its close elliptical satellites NGC 205 and NGC 221 form a triple system, and the farther satellites NGC 147 and NGC 185 can be added to produce a quintuple system. Among the nearest galaxies, NGC 147 and NGC 185 constitute a double system. Thus, more than half of the 19 nearest galaxies are components of double and multiple systems. Roughly the same proportion of stars in the Galaxy are multiple stars.

In 1811 W. Herschel studied 139 double objects, which he called double nebulae, and established great physical likeness between their components. Lundmark (1926) examined some 8000 NGC objects and found up to 200 double and multiple galaxies.

Currently the main catalogue of double and binary galaxies is Holmberg's catalogue (1937), which lists 827 objects. The selection criterion used by Holmberg was the inequality

$$\frac{\xi}{D_1+D_2} \ll 2, \qquad (60)$$

where D_1 and D_2 are the diameters of the galaxies, and ξ is the separation of their centers. This is a fairly rigid criterion and the number of optical

double galaxies in Holmberg's catalogue should be negligible. The catalogue lists galaxies up to 16^m, but it is comprehensive only up to $14^m.5$.

The frequency of systems of different multiplicities is the following:

Multiplicity	Number of systems
2	695
3	96
4	22
5	6
6	8

The ratio between systems of different multiplicities is approximately the same as for double and multiple stars.

It has been further established that early-type galaxies more frequently occur in close pairs, whereas late-type galaxies are found in wide pairs.

Ambartsumyan (1956) established that the number of double and multiple galaxies (like that of stars) is much greater than the number corresponding to dissociation equilibrium, i.e., to a state whereby the number of multiple systems forming through interaction between galaxies is equal to the number of systems which break up. Hence it follows that, at least in the neighborhood of our Galaxy, the predominant process is that of breakup of double and multiple systems.

A confirmation of this conclusion, in Ambartsumyan's opinion, is provided by the great number of Trapezium-type multiple galaxies, i.e., systems where all the three mutual separations of the components are comparable. Systems of this type are inherently unstable, since the motion of the components is non-Keplerian. Even if the total energy of a Trapezium system is negative, it will decay through internal interactions in a time which is only slightly greater than the time required for the components to traverse a path equal to the diameter of the system. Decay through internal interaction stops when the system is left only with two components; after that the double system may break up only following an encounter with other galaxies. Ambartsumyan found that among 132 multiple galaxies in Holmberg's catalogue, 87 have three-component configurations where the ratio of the maximum to the minimum mutual separation is at most 2.5. These systems can be safely classified as Trapezium type. Although their proportion out of the total number of multiple galaxies is exaggerated due to Holmberg's rigid selection criterion, Ambartsumyan found that Trapezium systems accounted at least for half of all the multiple systems. Their relative number is thus substantially greater than the relative number of Trapezium-type systems among multiple stars.

The existence of double and multiple galaxies in great numbers, and the high proportion of Trapezium systems among them, led Ambartsumyan to the conclusion that the component galaxies of these systems were of common and fairly recent origin.

An interesting feature of double and multiple galaxies is that many show lanes of luminous matter joining the different components. These lanes, arms or bridges, described by Zwicky (1956), are mainly observed in

close systems (Figure 217). Zwicky notes that each plate taken with the 48-in. Schmidt shows about a dozen double or triple systems whose components are joined by luminous matter. The bridges between galaxies are generally extensions of the spiral arms of these galaxies. Like the spiral arms, they are bluish in color. The nearby galaxy NGC 5194 (M51), which has been resolved into individual stars, shows that the bridges are made up of hot giants, gaseous and dust nebulae, and multitudes of lower luminosity stars. According to Vorontsov-Vel'yaminov (1958), the bridges joining E and S0 galaxies are short, wide, and bright, whereas Sb and Sc galaxies may be joined by long, narrow, and weakly luminous bridges. On the average, the longer the bridge, the narrower.

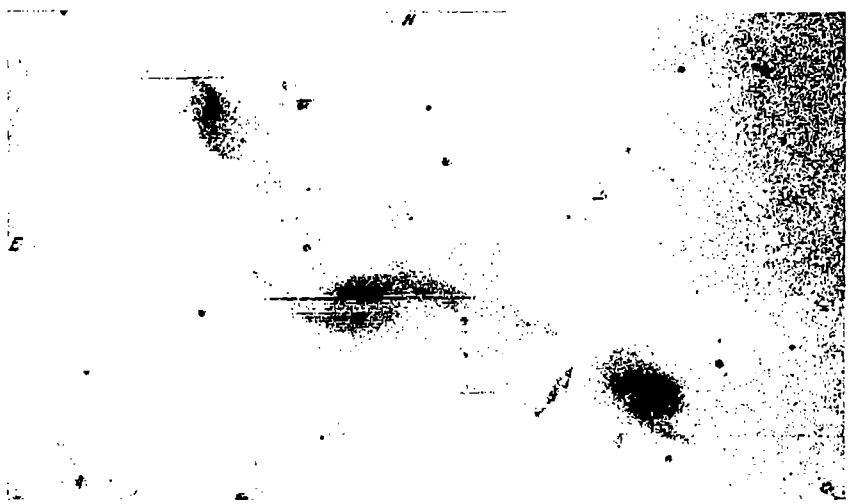


FIGURE 217

Alongside with bridges joining close galaxies, there are also instances of prominences protruding from one galaxy toward its companion, without actually joining the two. Vorontsov-Vel'yaminov calls these prominences tails. The tail projecting in the direction of the companion often has a counterpart which extends in the opposite direction. This is probably indicatory of a tidal interaction between galaxies.

There are currently two points of view on the nature of the bridges and the tails. According to Zwicky, these formations are the result of interaction between the close galaxies. Moreover, Zwicky is of the opinion that the spiral arms are the fragments of bridges which are severed by the rotation of the galaxies following an accidental encounter.

According to the other point of view (Ambartsumayn, 1958), the bridges are pulled out from the body of the system when two galaxies originating from a common core start receding from each other. The bridges are thus the "umbilical cord" linking the two separating galaxies into one.

Vorontsov-Vel'yaminov pointed to the following signs of interaction between close galaxies: 1) the facing parts of the galaxies are fainter, the

spiral structure in this region is greatly distorted or is absent altogether, 2) there are pairs or small groups of galaxies immersed in a common cloud of mist.

In 1959 Vorontsov-Vel'yaminov published Part I of his Atlas and Catalogue of Interacting Galaxies, which lists 355 systems.

§ 189. **Masses of double galaxies.** The determination of the masses of double galaxies is based on the assumption that the two components are physically linked and move one relative to the other in closed orbits. Near the perigalaxy the orbital velocities are thus greater than the circular velocity at that point, and at the apogalaxy the orbital velocity is conversely less than the corresponding circular velocity. Statistically, the motion can therefore be regarded as circular. Equating centrifugal and centripetal forces, we thus get

$$M_1 + M_2 \approx 2M = 1.12 \cdot 10^{-3} a v^2, \tag{61}$$

where M is the mass in units of solar mass, a is the radius of the relative orbit in astronomical units, v is the circular orbital velocity in km/sec.

Let Δv be the observed difference in the radial velocities of the components, Φ the angle between the circular velocity and the line of sight, ψ the angle between the radius-vector and the line of sight, s the angular separation of the components in minutes of arc, r the distance to the system in parsecs. Then

$$\Delta v = v \cos \Phi, \tag{62}$$

$$s = \frac{a \sin \psi}{60 r}. \tag{63}$$

Since v is perpendicular to a, we have

$$\left.\begin{aligned} \cos \Phi &= \sin i \cdot \cos \Omega, \\ \cos \psi &= \sin i \cdot \sin \Omega, \end{aligned}\right\} \tag{64}$$

where i is the inclination of the orbit to the tangent plane, Ω is the distance of the second component from the node of the relative orbit.

Combining Eqs. (61)—(63), we get

$$M \cdot f = M \sin \psi \cos^2 \Phi = 0.0336 \cdot r \cdot s \Delta v^2. \tag{65}$$

We can approximately take

$$\overline{M} = 0.336 \frac{\bar{r} \cdot \bar{s} \cdot \overline{\Delta v^2}}{\bar{f}}. \tag{66}$$

The average value of f is readily calculated from (64):

$$\bar{f} = \overline{\sin^2 i \cos^2 \Omega \sqrt{1 - \sin^2 i \sin^2 \Omega}} = \frac{\int_0^{\frac{\pi}{2}} d\Omega \int_0^{\frac{\pi}{2}} \sqrt{1 - \sin^2 i \sin^2 \Omega} \cos^2 \Omega \sin^3 i \, di}{\int_0^{\frac{\pi}{2}} d\Omega \int_0^{\frac{\pi}{2}} \sin i \, di} \cong 0.295.$$

Thus,

$$\overline{M} = 0.114 \bar{r} \bar{s} \overline{\Delta v^2}. \tag{67}$$

Eq. (67) enables us to estimate the mean mass of double galaxies if the radial velocities of the components are known and the distances have been determined by some method.

Page (1952) obtained $\overline{M} = 8 \cdot 10^{10} M_\odot$ from the radial velocities of the components of 20 double galaxies. The mean error is estimated at $3 \cdot 10^{10} M_\odot$. The relative radial velocities of the binary components in Page's measurements reach 640 km/sec.

In 1959, Page and Thornton repeated this study on a larger sample (65 double galaxies). According to Eq. (67) the mean mass is proportional to the distance and hence inversely proportional to Hubble's constant (if the distances are determined from radial velocities). Taking 100 km/sec · Mpc^{-1} for Hubble's constant, Page and Thornton obtained $2.5 \cdot 10^{11} M_\odot$ for the mean mass of a galaxy. For 45 E and S0 galaxies separated from the total sample, the mean mass was found to be $4 \cdot 10^{11} M_\odot$, and for the 20 S and I galaxies it was $0.3 \cdot 10^{11} M_\odot$.

Note that if there is no physical bond between the system components, Eq. (61) is replaced by the inequality

$$2M < 1.12 \cdot 10^{-3} a v^2$$

and (67) by

$$\overline{M} < 0.114 \overline{rs} \overline{\Delta v^2}. \tag{68}$$

Thus, if our fundamental assumption is in error, the masses obtained from (67) are too high.

§ 190. **Clusters of galaxies.** The distribution of galaxies in space is far from being uniform. They group into aggregates of various types and sizes. Whereas only a minor proportion of stars occur in star clusters, the majority being stars of the general galactic field, galaxies are predominantly observed in condensations.

The spatial distribution of galaxies was studied by Zwicky. He distinguishes between aggregates of different kinds (1957):

1) Clusters of galaxies. These formations contain hundreds and even thousands of objects. A highly important feature of a cluster is the presence of one or several concentration centers for its members. Some clusters show signs of regular structure. The density of galaxies in the central parts of a cluster is thousands and even millions of times greater than the field density. Clusters with circular symmetry are called regular. All the rest are irregular clusters.

2) Groups of galaxies containing less than a hundred objects. The density of galaxies in groups is substantially higher than the field density. Groups show no marked concentration toward the center.

3) Clouds of galaxies. These are large groups with hundreds and thousands of galaxies. The clouds are irregular, without any central concentration.

4) Clouds of groups. The density of galaxies in the space between the component groups in these clouds is higher than the metagalactic field density.

5) Clouds of clusters, the largest among the observed condensations; double, triple, and multiple clusters are observed. Large clouds may contain many tens of clusters, with over 100,000 galaxies.

Condensations of galaxies are studied and identified by counting the number of galaxies in selected sky areas.

Clusters of galaxies are the most remarkable objects in this classification.

We will describe Zwicky's results obtained for the Coma Berenices cluster. This cluster has a distinct concentration center and circular symmetry. Zwicky counted the galaxies in a series of rings concentric with the cluster using a 16-in. Schmidt, which gives a limiting magnitude of $16^m.5$, and a 48-in. Schmidt, which reaches $19^m.0$ apparent magnitude.

The distance of this cluster in Sandage's scale is 85 Mpc, and the distance modulus is $34^m.7$. Thus, the 16-in. Schmidt counts galaxies brighter than absolute magnitude $-18^m.2$, and the 48-in. Schmidt reveals galaxies brighter than $-15^m.7$. The brightest galaxies of this cluster have apparent magnitude $+13^m.2$, which corresponds to absolute magnitude $-21^m.5$. That the brightest galaxies of this cluster are brighter than our Galaxy or NGC 224 is not surprising and it provides a further argument in favor of Sandage's scale. The counts show that the number of galaxies of a given magnitude increases monotonically with apparent magnitude, i.e., the luminosity function monotonically increases from $M = -21^m.5$ to $-15^m.7$.

TABLE 16

Ring boundaries in minutes	Number of galaxies up to $16^m.5$	Number of galaxies up to $19^m.0$	Number of galaxies up to $16^m.0$ per 1 sq. deg	Number of galaxies up to $19^m.0$ per 1 sq. deg	n_1	n_2	n_2-n_1	$\frac{n_2-n_1}{n_1}$
0—5	31	60	1420	2716	1413	2546	1133	0.820
5—10	42	125	641	1910	634	1740	1106	1.75
10—15	53	152	485	1394	478	1224	746	1.56
15—20	44	166	288	1086	281	916	635	2.26
20—30	87	325	200	745	193	575	382	1.98
30—40	66	365	108	600	101	430	329	3.26
40—50	72	395	91	502	84	332	248	2.96
50—60	73	456	76	475	69	305	236	3.42
60—70	69	453	61	400	54	230	176	3.3
70—80	34	471	26	359	19	189	170	8.9
80—90	32	441	22	298	15	128	113	7.5
90—100	41	530	25	320	18	150	132	7.4
100—120	61	1091	16	291	9	121	112	13
120—140	49	1311	11	289	4	119	115	29
140—160	50	1347	10	257	3	87	84	28
160—180		1579		266		96		
180—200		1828		275		105		
200—220		2100		287		117		
220—240		2233		278		108		
240—270		3436		257		87		
270—300		3466		232		62		
300—330		3629		220		50		
330—360		3992		221		51		

The results of these counts are listed in column 2 and 3 of Table 16. The next two columns give the count per 1 sq. degree of the sky, obtained by dividing the counts in columns 2 and 3 by the respective ring areas.

Extraneous galaxies lying nearer or farther than the cluster are accidentally projected onto the cluster area. To correct for this accidental projection effect, the average number of galaxies up to $16^m.5$ and up to $19^m.0$ per 1 sq. degree outside the cluster area at the same latitude should be subtracted from the figures in columns 4 and 5. Zwicky found that the average number of galaxies up to $16^m.5$ and $19^m.0$ in 1 sq. deg of the field was 7.4 and 170, respectively. After subtraction, columns 6 and 7 give the numbers n_1 and n_2 of cluster galaxies up to $16^m.5$ and $19^m.0$ in 1 sq. deg of the sky for the different concentric rings.

Column 8 gives the number of weak galaxies with apparent magnitudes from $16^m.5$ to $19^m.0$ per 1 sq. deg in the cluster. Column 9 gives the ratio of the number of weak galaxies to the number of bright galaxies.

The data in the table show that the bright galaxies fix the boundary of the cluster at a distance of some 160' from the center. If, however, the weaker galaxies are also considered, the cluster is found to have a radius of over 360'. The ratio of the number of galaxies per 1 sq. deg at the center of the cluster to the number of galaxies per 1 sq. deg of the metagalactic field is 200 for the bright galaxies, and only 7 for the weak galaxies. The bright galaxies thus show a much more pronounced central concentration, and their proportion out of the total number of galaxies steadily and fairly rapidly decreases with distance from the center. If we change over from surface to spatial distribution (as for globular star clusters, see §173), the proportion of the bright galaxies will diminish even faster away from the center.

Among the nearest galaxies, higher mass is an indication of higher luminosity. This relation apparently holds true for galaxies in clusters also. If this is so, the massive galaxies in the Coma Berenices cluster seem to be nearer the center, whereas low-mass galaxies also occur at the periphery. This effect was attributed by Zwicky to the irregular gravitational field in the cluster. This conclusion is further supported by the almost spherical distribution of galaxies in the cluster.

It should be noted, however, that Zwicky's interpretation is not the only one possible. Another possibility is that the cluster of galaxies arose from some primeval formation with perfect spherical symmetry and density concentration toward the center. Then massive galaxies would naturally form in the central regions, where the density of matter was highest.

The linear diameter of the Coma Berenices cluster is 8 Mpc from the bright galaxies and 18 Mpc from the weak galaxies.

The total number of galaxies up to $16^m.5$ in the cluster is 654, and of galaxies up to $19^m.0$, 10,724. This gives an average density of 3.5 galaxies per 1 Mpc^3. In the central regions, there are over 100 galaxies per 1 Mpc^3, and the distances between galaxies are comparable to the size of the individual galaxies.

The composition of galaxies in Coma Berenices is somewhat unusual. The cluster contains absolutely no galaxies with prominent spiral structure, but besides the E galaxies there is an abundance of S0 galaxies. A similar situation is observed in another regular and centrally dense cluster of galaxies, the one in Corona Borealis. We recall that S0 galaxies are exceedingly flattened and have a nucleus, like the spiral galaxies, but they show no spiral arms and no traces of dark matter. The last two features are intimately interrelated, since the spiral arms are formed by hot giants, and hot giants are genetically linked with dark matter. Proceeding from

these considerations, Spitzer and Baade (1951) suggested that the S0 galaxies in dense clusters are the product of collisions between spiral galaxies. In a head-on collision between two spiral galaxies, the star of each galaxy will sweep through the other almost unimpeded because of their enormous free paths. The overall deformation of the colliding systems will also be insignificant. The diffuse matter in these galaxies, however, will effectively collide, since the free path of the diffuse cloud particles is small compared to the size of the galaxies. The diffuse matter will thus be swept out from the star system, and a normal S galaxy will develop into S0. This hypothesis presupposes that the hot giants are genetically linked with diffuse matter in such a way that they vanish when the diffuse matter is removed from the galaxy. Other, less dense clusters (we are concerned with the partial pressure of S and S0), e.g., the Virgo cluster, contain S galaxies, and S0 galaxies are fairly rare. Spitzer and Baade explain this by the low probability of collisions in the low-density cluster, so that only few of the spiral galaxies have so far lost their diffuse matter.

Table 17 lists Zwicky's general data on three regular clusters of galaxies.

TABLE 17

Characteristic	Coma Berenices cluster	Cancer cluster	Pegasus cluster
Galactic longitude	26°	170°	58°
Galactic latitude	+87°	+30°	−49°
Distance in megaparsecs . . .	85	55	45
Diameter in degrees according to galaxies up to $19^m.0$ . .	12.0	7.0	2.0
Absolute magnitude of the brightest member	−21.5	−20.0	−20.8
Number of galaxies up to $16^m.5$	654	84	62
Number of galaxies up to $19^m.0$	10724	300	369
Average density of galaxies per 1 Mpc^3	3.5	1.9	180

Note the differences in the population and the spatial density of these three regular clusters.

The average density of galaxies in the metagalactic field is about 0.05 per 1 Mpc^3. The density in clusters is thus much greater than the metagalactic field density.

An example of an irregular cluster of galaxies is the Virgo cluster. This is the nearest of the large clusters. Its distance modulus is $30^m.7$, and the distance is about 12 Mpc. The bright galaxies in this cluster occupy the area $250° < l < 270°$, $+50° < b < +70°$. Of the 34 galaxies listed in Messier's catalogue, 16 occur in the Virgo cluster. The Shapley—Ames catalogue includes 97 galaxies with $m \leqslant 12^m.3$ from this cluster and 106 galaxies with $12^m.3 < m < 13^m$. It contains thousands of weaker galaxies. Zwicky's photographs taken with the 48-in. Schmidt show a great number

of weak irregular galaxies of low surface brightness, i.e., III galaxies. The size of the cluster increases as we pass to weaker galaxies, as in the case of regular clusters. The overall linear dimensions of this enormous cluster are approximately 8 Mpc and the mean density in the cluster is 15 galaxies per Mpc^3. The bright galaxies of this cluster include numerous giant elliptical galaxies which show a more pronounced central concentration than the spirals. In Zwicky's opinion this may be due to the higher mass of E galaxies compared to S galaxies for constant luminosity.

The Virgo cluster has a definite central concentration. This concentration, however, is less pronounced than in the Coma Berenices cluster. In general, irregular clusters show a much less pronounced central concentration than the regular clusters do.

The irregular clusters also include those in Ursa Major, Centaurus, Fornax, Leo.

The best known group of galaxies is the so-called Local Group, which includes our star system. This group comprises the first 19 objects in Table 3: these are six spirals, six elliptical galaxies, and seven irregular galaxies, The linear dimensions of the Local Group are about 0.8 Mpc. The density of galaxies in the Local Group is thus approximately 40 per Mpc^3. When comparing this density with the previously mentioned figures for clusters and the metagalactic field, we must remember that in the Local Group even the faintest members — dwarfs with absolute magnitudes between -10^m and -12^m — are counted. In the metagalactic fields and other clusters these weak objects are simply undetectable. The only exception is the Virgo cluster, where Zwicky observed galaxies of similar magnitude. The farther the cluster, the more limited are our possibilities concerning the detection of the weak galaxies, and this explains the pronounced selection. The density in the far clusters and in the metagalactic field is thus lower than the true figure.

Other groups are the M81 and the M101 group listed in Table 2. Clearly a substantial proportion of galaxies in these groups have not been discovered yet. The absence of elliptical galaxies in these groups and of irregular dwarf galaxies in M101 is highly suspicious.

The Local Group together with the groups M81 and M101 form a small cloud of groups in Zwicky's classification.

In 1953 Vaucouleurs investigated the distribution of the bright galaxies in the Shapley — Ames catalogue over the sky and established that these galaxies (the members of the Local Group excluded) concentrate near the great circle which is precisely perpendicular to the galactic equator, between galactic longitudes of 105 and 285°. In the northern galactic hemisphere, the concentration circle passes through the Virgo cluster. If we ignore the dense group of galaxies localized in a small part of the southern sky, a strip some 12° wide around the concentration circle, which is a mere 10% of the total sky area, contains about 2/3 of all the galaxies brighter than 12^m. If we take into consideration the avoidance zone associated with interstellar absorption of light, we obtain in this strip a surface density of galaxies 10 times higher than over the rest of the sky.

Vaucouleurs considered this fact as evidence in favor of the existence of a large disk-shaped condensation of galaxies, which he called the Local Supersystem. As we pass to ever weaker galaxies, the concentration diminishes, unlike the galactic concentration of stars, This is apparently

associated with the fairly small size of the Local Supersystem, which does not include the distant galaxies.

Having investigated the density of galaxies along the concentration circle, which is the equator of the Local Supersystem, Vaucouleurs came to the conclusion that the center of the Local Supersystem lies in the Virgo cluster, which can be considered as the nucleus of the Local Supersystem. The Local Supersystem is highly flattened (degree of flattening 8), which points to pronounced rotation effects. Its diameter is some 20—30 Mpc (on Sandage's scale), and the distance from the Galaxy to the rim is 2—4 Mpc.

Thus, slightly deviating from Vaucouleurs's terminology, we can interpret his results as indicating that the Virgo cluster is highly flattened and larger than was previously assumed. Our Galaxy lies in the equatorial plane of this cluster, just at its rim. The Local Group is one of the peripheral condensations of galaxies in the Virgo cluster.

If we remember that the Virgo cluster is much nearer than any other large cluster, Vaucouleurs's conclusions seem quite probable.

According to Vorontsov-Vel'yaminov, the concentration of galaxies toward the equator of the Local Supersystem holds true at least for galaxies of up to 18^m.

The dense group of galaxies localized in a small part of the southern sky is regarded by Vaucouleurs as a neighbor supersystem.

The concentration of optically visible galaxies near the plane of the Local Supersystem was confirmed by Kraus and Ko (1953), Brown and Hazard (1953), and Kraus (1954), who showed that the radio emission also concentrated in that plane.

This is a highly important result, since the radio emission is not obstructed by the dark matter of our Galaxy. The concentration of radio sources even proved to be greater than that predicted from the optical concentration. In Shklovskii's opinion (1954), this may be due to the fact that a substantial fraction of the radio waves originates in the diffuse matter concentrated in the plane of the Local Supersystem.

Shakeshaft and Baldwin investigated the radio intensity at 1.9-m wavelength along the concentration circle and failed to detect any correlation with the number of bright galaxies. This result is understandable if almost the entire radio power is emitted by the diffuse matter in the plane of the Local Supersystem and the contribution from individual galaxies is insignificant. The authors themselves suggest that the observed radio emission is of galactic origin.

A radio survey at 3-m wavelength (100 MHz) carried out by Bolton and Westfold (1950) did not reveal any enhancement of radio intensity in the part of the southern sky which Vaucouleurs identified with a neighbor supersystem.

Wider horizons for statistical studies of clusters are now being opened in connection with the compilation of the relevant catalogues.

Abell (1958) compiled a catalogue of 1712 rich clusters of galaxies located using the Palomar Sky Atlas. A fairly homogeneous statistical sample of 1682 clusters was selected from among these clusters, satisfying the following criteria:

1. Each cluster should contain a minimum of 50 galaxies with magnitudes at most 2^m higher than the magnitude of the third brightest member in the cluster.

2. The fifty brightest members of the cluster should lie within a circle of radius $\frac{4.6 \cdot 10^5}{c \frac{d\lambda}{\lambda}}$ mm on plates taken with the 48-in. Schmidt (c is expressed in km/sec).

3. The red shift of the cluster should fall between 6000 and 60,000 km/sec.

4. The cluster should be located at a sufficiently high galactic latitude, to avoid the effect of galactic absorption.

The frequency $N(n)$ of the cluster populations rapidly falls off with increasing n (note that $n > 50$). This distribution led Abell to the conclusion that the average number of galaxies in a cluster is markedly less than 50. This result is in good agreement with the results of the theory of fluctuations, with allows for the patchy structure of the dark matter (see § 188).

Herzog, Wild, and Zwicky are now working on a catalogue of rich clusters for the sky north of $\delta = -30°$ using the 48-in. Schmidt. The surveyed part of the sky covers 30,940 sq. deg, some 10,000 of which are obstructed by dark matter. The remaining 21,000 sq. deg contain about 10,000 rich clusters. A rich cluster is one containing at least 50 galaxies with apparent magnitudes between $[m_{max}, m_{max} + 3]$, where m_{max} is the apparent magnitude of the brightest galaxy. Condensations with surface density of galaxies exceeding by more than a factor of 10 the surface density of field galaxies were also classified as clusters.

This catalogue uses the following classification of clusters: 1) compact clusters are those with a single distinct point of concentration; the surface density of galaxies near this point is such that they actually (optically) touch one another. Most of the compact clusters are spherically symmetric. 2) Semicompact clusters are those where the galaxies around the point of maximum concentration occur at separations several times greater than their size, or else the cluster has several concentration points. 3) Open clusters are those without any distinct concentration points.

According to this classification, the Coma Berenices cluster is compact, the Virgo cluster is semicompact, and the Ursa Major cluster is open.

Zwicky (1955) counted the galaxies in the identified clusters using photographs taken with the 48-in. Schmidt in a sky area covering 40 sq. deg. There were 921 clusters in this area, containing 118,191 galaxies, which gives an average of 128 galaxies per cluster. Note, however, that the average number of galaxies obtained in this way is definitely overestimated, since small clusters lose their individuality due to projection effects, and only the large clusters stand out distinctly. Each photographic plate showed the following number of clusters: 41, 53, 54, 34, 64, 36, 45, 35, 30, 15, 22, 9, 17, 56, 11, 24, 20, 8, 34, 24, 36, 54, 41, 64, 55, 39. The observed fluctuations in the number of clusters per plate are much greater than the natural fluctuations in case of uniform distribution. They are either due to fluctuations in interstellar absorption or to a tendency of clusters to aggregate into clouds. Both factors apparently are significant. Anyhow, Zwicky's data show that the hypothesis of uniform distribution of clusters and zero fluctuations in interstellar absorption does not fit the observations.

§ 191. Determination of galactic masses in clusters. The mass—luminosity relation in star systems. Clusters of galaxies provide an alternative method for the determination of galactic masses. The following information is required:

Also

$$N = 2\int_0^R S(x)\,dx. \tag{76}$$

Inserting (75) and (76) in (70), we get

$$m = \frac{\overline{v^2}\int_0^R S(x)\,dx}{G\int_0^R S^2(x)\,dx}. \tag{77}$$

If the strip width is now measured in angular units β, we have

$$x = r\cdot\beta,$$
$$S(x)\,dx = s(\beta)\,d\beta,$$
$$\int_0^R S(x)\,dx = \int_0^\alpha s(\beta)\,d\beta,$$
$$\int_0^R S^2(x)\,dx = \frac{1}{r}\int_0^\alpha s^2(\beta)\,d\beta,$$

where α is the angular radius of the cluster in radians, and r is the distance to the cluster.

Eq. (77) thus takes the form

$$m = \frac{\overline{v^2}r\int_0^\alpha s(\beta)\,d\beta}{G\int_0^\alpha s^2(\beta)\,d\beta}. \tag{78}$$

This is the working formula for the calculation of the mean galactic mass in the cluster.

According to Holmberg, the mean square residual velocity in the Coma Berenices cluster (as measured for 22 galaxies in the cluster) is 1430 km/sec. The integrals on the right in (78) were evaluated by Schwarzschild using Zwicky's counts of galaxies in the Coma Berenices cluster. If we take 85 Mpc for the distance to the cluster, we get

$$m = 2.6\cdot 10^{45}\,\mathrm{g} = 1.3\cdot 10^{12} M_\odot. \tag{79}$$

The stellar population of galaxies is conveniently characterized by the ratio f of the galactic mass to luminosity. It is generally expressed in units of the corresponding ratio for the Sun. The values of f for various objects were assembled by Schwarzschild (1954) (Table 18).

Table 18 shows that f increases in late-type galaxies, where the mass per unit luminosity is greater. This is quite understandable, since late-type galaxies are mainly Population II systems, and their constituent stars are characterized by high values of f.

TABLE 18

Object	Distance, kpc	Type	Luminosity (solar units)	Mass (solar units)	f
Sun's neighborhood		Sb or Sc			4
NGC 598	480	Sc	$1.4 \cdot 10^9$	$5 \cdot 10^9$	4
LMC	44	II	$1.2 \cdot 10^9$	$2 \cdot 10^9$	2
NGC 224	460	Sb	$9 \cdot 10^9$	$1.4 \cdot 10^{11}$	16
NGC 3115	2,100	E	$9 \cdot 10^8$	$9 \cdot 10^{10}$	100
NGC 221	460	E	$1.1 \cdot 10^8$	$2.5 \cdot 10^{10}$	200
Average S and I in double galaxies		S, I	$2.4 \cdot 10^{10}$	$3 \cdot 10^{10}$	1.3
Average E and S0 in double galaxies		E, S0	10^{10}	$4.0 \cdot 10^{11}$	41
Average galaxy in Coma Berenices	85,000	E, S0	$5.5 \cdot 10^9$	$1.3 \cdot 10^{12}$	240

§ 192. **Radio emission from galaxies.** The distribution of discrete radio sources over the sky shows that these sources are divided into two distinct groups. Sources of one group are confined in a narrow belt near the galactic equator, showing pronounced galactic concentration. The second group comprises sources which lie outside the equatorial plane and are distributed at random, without any signs of galactic concentration.

The radio sources of the first group, as expected, are galactic objects and some of them have been identified with gaseous nebulae and remnants of supernovae and novae.

The radio sources of the second group are either exceedingly near stars (very near stars do not show any galactic concentration) or distant extragalactic objects. None of these sources have so far been identified with any individual star. If we assume that these are optically weak, invisible, near stars, we will have to revise our notions of the stellar density around the Sun, despite their excellent fit with general dynamic considerations. The radio sources of the second group were thus regarded from the start as probable extraglactic objects. This conclusion was finally confirmed in 1949 when Bolton, Stanley, and Slee identified two discrete sources with the galaxies NGC 5128 and NGC 4486. Soon after that many more identifications were made, and we are now fairly certain that the discrete radio sources without galactic concentration are in fact galaxies or clusters of galaxies.

For purposes of comparison between the radio and the optical radiation power of discrete sources, one generally defines the radio magnitude of the source. In Brown and Hazard's system it is defined by

$$m_r = -53.4 - 2.5 \lg S(158), \tag{80}$$

where $S(158)$ is the flux density at frequency of 158 MHz. Its dimensions are watt $\cdot$ m^{-2}Hz^{-1}. The radio to optical intensity ratio of a source, which we call the relative radio brightness, is then defined by the difference

$$m_r - m_p, \tag{81}$$

where m_p is the apparent photographic magnitude of the object. The smaller the difference (81), the greater is the relative radio brightness of the source. $m_r - m_p$ slightly depends on the distance of the galaxy. Since the radio intensity falls off more slowly with increasing wavelength than the optical intensity does, the red shift of the distant galaxies should reduce $m_r - m_p$, and hence increase the relative radio brightness.

In terms of relative radio brightness, the galaxies are divided into two distinct groups. For the great majority of galaxies, including all the nearest galaxies, the relative radio brightness is small. Their $m_r - m_p$ is greater than -2^m. These are the normal galaxies. The relative radio brightness is particularly small for E, S0, and III galaxies. None of the brightest galaxies of these types emit any detectable radio waves. Since the threshold magnitude m_r for the modern radio telescopes is about $+10^m$, and the brightest E, S0, and III galaxies have apparent magnitudes of $+8^m.9$, $+9^m.1$, $+9^m.3$, $+9^m.5$, their $m_r - m_p$ corrected for light absorption should be over $+2^m$ or even $+3^m$.

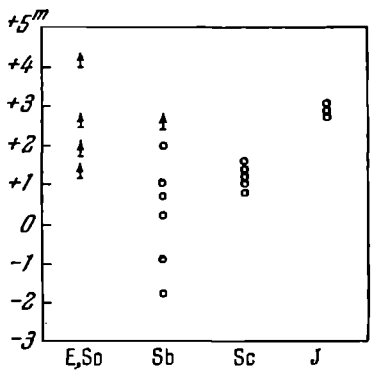


FIGURE 218

Minkowski (1958) constructed from Brown and Hazard's data and also from Mills's finding the diagram of $m_r - m_p$ for bright galaxies of various types (Figure 218). The arrow pointing up shows that the radio power of the particular galaxy is undetectable and $m_r - m_p$ is greater than the value of this difference at the base of the arrow.

We see from the diagram that III galaxies are weak radio sources and the dispersion of their $m_r - m_p$ value is very small. The dispersion of $m_r - m_p$ is also small for Sc galaxies, but their relative radio brightness is somewhat higher. In Sb galaxies, $m_r - m_p$ show considerable disperison, ranging from -2^m to $+2^m$. The is apparently associated with the high spread of the spectra of Sb galaxies (§ 179).

The absence of noticeable radio emission from E, S0 and III galaxies, combined with the complete lack of radio power from globular clusters in the Galaxy, shows that Population II stars are not active at radio wavelengths.

The $m_r - m_p$ of spirals on the average decreases with decreasing inclination i. This is explained by the stronger optical screening of the galaxy by the dark matter at higher i, whereas at radio wavelengths the dark matter is transparent. The inclination correction to $m_r - m_p$ is quite substantial (up to $+2^m$). The correction for i, however, did not markedly reduce the dispersion of the $m_r - m_p$ values for Sb galaxies. They remained spread between $-0^m.3$ and $+3^m.6$.

Minkowski links the radio emission with the presence of spiral arms. II are weak radio sources probably because their spiral structure is insufficiently developed. It seems that radio emission and spiral structure require primarily large quantities of diffuse matter and also sufficiently

strong magnetic fields. The Large Magellanic Cloud is rich in diffuse matter. Its magnetic field, however, is probably very weak, as its radio emission is insignificant and its spiral arms are inconspicuous.

Of the greatest interest is the second group of galaxies characterized by enhanced relative radio brightness. These radio galaxies, or peculiar galaxies, constitute a negligible fraction out of the total galactic population, but their intense radio emission clearly separates them from the multitudes of normal galaxies.

For the radio galaxies $m_r - m_p$ reaches large negative values.

Table 19 compiled by Minkowski lists the identified radio galaxies.

TABLE 19

No.	Object	$m_r - m_p$	Apparent characteristics	Spectrum
1	NGC 1275	−6	d, p	pe
2	NGC 1316	−3	p?	n
3	NGC 4038-9	−1	d, p	ne
4	NGC 4486	−5	p	pe
5	NGC 5128	−3	d, p	pe
6	Cygnus A	−14	d, p	pe
7	Hydra A	−12	d	pe?
8	Hercules A	−12	d, p	pe?
9	NGC 2623	−5	p	n
10	NGC 4782-3	−3	d	n
11	NGC 6166	−7	d	ne

Under apparent characteristics, d denotes a double galaxy, p a peculiar aspect, e.g., ejections tails. Under spectrum, pe points to spectral peculiarities, n normal spectrum, ne normal spectrum with emission.

The first six galaxies on the list were firmly indentified with radio sources, the seventh and the eighth are less certain, and the last three are fairly uncertain. The identification difficulties are associated with the extreme optical weakness of most radio galaxies; moreover, the position of the discrete radio sources cannot be pinpointed with high accuracy. Therefore, the fairly large area in which the radio source is localized (an area of some 100 sq. min) may contain several galaxies of comparable brightness. In this case, additional argument in favor of the identification is provided by the peculiar aspect and the peculiar spectrum of the optical galaxy.

Cygnus A — a 18^m galaxy ($+16^m$ with correction for absorption) — is the most remarkable object on the list. Here the identification was made possible by a most pronounced peculiarity of the object. This is an exceedingly close double galaxy with a most peculiar spectrum. Although optically Cygnus A is a weak object (because of its low brightness it was not included in NGC and is designated Cygnus A, meaning that it is the strongest discrete radio source in the constellation of Cygnus), it is a most intense radio source, with $m_r = +2^m$, so that it has a tremendous relative radio brightness $m_r - m_p = -14^m$. Many weak radio sources possibly remain without optical counterparts because these are also objects of the Cygnus A type. Such an object with $m_r = 8^m$ has $m_p = 22^m$, and there is virtually no

hope of reliable identification, since the radio source area contains a great number of 22-mag. galaxies and the optical peculiarities, if any, of such weak objects cannot be distinguished. Cygnus A objects of radio magnitude $m_r = +10^m$ have $m_p = +24^m$, and are thus entirely inaccessible to optical observations. Thus, if we assume the existence of Cygnus A objects among radio sources with $m_r = +10^m$, we can safely state that already now the radius of the Universe covered by radio observations is much greater than the radius of the optical Universe.

While the applicability of radio techniques to the normal galaxies is limited by their weak radio emission, the studies of radio galaxies are naturally limited by their extreme optical faintness.

Baade and Minkowski (1954) interpret Cygnus A as galaxies in collision—a pair of late-type spirals. If the galaxies indeed collide, the diffuse matter of the two components is intermixed and the optical spectrum is expected to show certain peculiarities. Indeed, the spectra taken with the 200-in. and the 100-in. telescopes identify the double galaxy Cygnus A as a peculiar object. It shows bright forbidden lines. More than 50% of its radiation is concentrated in emission lines. The red shift gives 200 Mpc for its distance. The total power radiated at radio wavelengths can be estimated at $3 \cdot 10^{44}$ erg/sec, whereas the optical power is about $2 \cdot 10^{44}$ erg/sec.

The powerful radio emission originates in the colliding diffuse clouds of the two galaxies. Shklovskii (1954) suggests that when diffuse masses collide, a substantial fraction of their kinetic energy is converted into the energy of relatively few relativistic particles, i.e., particles moving with tremendous velocities. These relativistic particles decelerate in the magnetic fields and emit at radio wavelengths. According to Baade and Minkowski, the radio galaxy NGC 5128 is the result of a collision of a spiral and an elliptical galaxy.

The radio galaxy NGC 4486 is definitely a solitary object, so that there can be no question of galaxies in collision. This galaxy, however, has a unique feature, a bright "prominence" 20" long and 2" wide in the nuclear region. The prominence has a continuous spectrum, without any absorption or emission lines, which fits the radiation of relativistic electrons moving in magnetic fields. Shklovskii suggests that the bright prominence and the powerful radio emission of NGC 4486 are the result of intragalactic collisions between large masses of interstellar gas. These gas aggregates (a kind of globular clusters, but containing interstellar gas) are observed in large numbers in NGC 4486.

Ambartsumyan has developed a different approach to the entire problem of radio galaxies (1956). He notes that whereas the total area of the radio source in Cygnus A is about 2' × 1', the separation between the optical nuclei of the two galaxies is about two seconds of arc. In this case, following the hypothesis of Baade and Minkowski, it would seem that the galaxies are almost in head-on collision. Offside collisions, however, should be much more frequent, and in principle they should lead to the same effects, though of lower intensity. Therefore, if Cygnus A are galaxies in collision, there should be an additional large number of similar objects of lower relative radio brightness, including some nearer objects with higher absolute radio intensity. There are, however, no such known objects.

Assuming the collision hypothesis, NGC 5128 must be regarded as a case of head-on collision, a highly improbable event. Ambartsumyan maintains that radio galaxies are the result of fission of a parent object into two fragments, two receding galaxies. The fission is accompanied by various processes which emit high-intensity radio waves. The closeness of the components in double radio galaxies is naturally explained by the fission hypothesis. It should be noted, however, that we are not familiar with any analogs of this tremendous fission process, and therefore the actual mechanism of fission and the attendant effects are still a subject of pure conjecture. According to Ambartsumyan, the prominence in NGC 4486 is also a result of fission in the nuclear region of this galaxy. An additional argument in favor of Ambartsumyan's hypothesis is that all the radio galaxies are supergiants in terms of optical luminosity. If fission should occur, then naturally we expect it to occur in supergiant objects. Within the framework of the collision hypothesis, on the other hand, we have to assume that the luminosity of galaxies increases during collision, because of the interaction between stars and diffuse matter, say.

Ambartsumyan's interpretation of radio galaxies is closely linked with his approach to the problem of multiple galaxies, discussed in § 189.

To conclude our discussion of radio galaxies, note that the identified radio galaxies are indeed peculiar optical objects. There are, however, similar peculiar optical objects without any signs of intense radio emission. Examples of such objects were cited by Vorontsov-Vel'yaminov and Vaucouleurs. This introduces an additional element of doubt into the applicability of the collision hypothesis to radio galaxies. It is incompatible with the fission hypothesis either.

In 1953 Brown and Hazard tried to detect the radio emission from clusters of galaxies situated at distances over which the radio emission of solitary normal galaxies is definitely undetectable. They recorded the radio emission of the clusters in Perseus and Ursa Major, and in both cases it was found to be stronger than expected from the total number of galaxies in the cluster (assuming normal galaxies). The authors thus concluded that these clusters contain radio galaxies, as well as normal galaxies. There are certain indications that NGC 1275 in the Perseus cluster is indeed a radio galaxy.

Of considerable interest is the frequency of the radio magnitudes of the known radio sources. The arbitrary constant k in the expression

$$m_r = k - 2.5 \lg H, \tag{82}$$

where H is the flux density in watt $\cdot$ m^{-2} Hz^{-1}, was chosen so that the number of objects up to a given radio magnitude was comparable with the number of optical objects (stars) up to the corresponding apparent magnitude. In this system of radio magnitudes the parameters are clearly different from those used in system (80).

Kraus (1955) published a table (Table 20) of the number of radio sources up to a given radio magnitude, based on the findings of a number of authors. For comparison, the last column gives the number $N(m)$ of optical objects (stars) up to given magnitude m.

We see that first the $N(m)$ of radio sources increases more slowly than that of the optical objects, but eventually the results correspond to a uniform distribution in space with $\frac{N(m+1)}{N(m)} = 3.98$. It thus seems that radio

sources are distributed relatively uniformly in space and their detection is independent of absorption. The slow increase in $N(m)$ for $m \leqslant 2$ seems to indicate that galactic objects prevail among the bright radio sources, while most of the weak radio sources are extragalactic objects.

TABLE 20

Magnitudes	$N(m_r)$				$N(m_p)$
	Kraus, Ko, and Matt (1954)	Bolton, Stanley, and Slee (1954)	Mills (1952)	Ryle, Smith, and Elsmore (1950)	Duncan (1926)
1	7	5	6	6	11
2	8	6	16	8	40
3	35	15	56	32	135
4	163	81			450
5	1000 ± 250	—	—	—	1500
6	—	—	—	—	4800
Balance	44	23	21	18	

The radio source counts carried out by Kraus, Ko, and Matt covered about 2/3 of the entire sky area. The number of radio sources up to 5th radio mag. can therefore be estimated at 1500.

§ 193. **The Metagalaxy. The structure of the Universe.** Like stars which collect into star clusters and star clouds, galaxies also occur in clusters and clouds. For stars, however, there are higher-order formations — star systems or galaxies, which differ from star clusters and clouds in their greater autonomy, closure, and independence of exterior objects (note that this property is also characteristic of globular clusters, which can be treated as miniature galaxies). It is for this reason that galaxies, and not star clusters and star clouds, are regarded as the next higher formation in the Universe after stars. The existence of star systems led to the idea of a gigantic system of galaxies possessing greater autonomy that the observed clusters and clouds of galaxies. This system by definition combines all the observed galaxies and is in a way the equivalent of a galaxy for stars on a much higher level. This hypothetical system of galaxies is called the Metagalaxy. It is assumed that the entire observable part of the Universe is the interior of the Metagalaxy. At present we simply have no methods for identifying extrametagalactic objects, even if such are observed.

The hypothetical part in the concept of the Metagalaxy is not the existence of a gigantic system of galaxies, which includes our Galaxy (a suitable realistic example is provided by Vaucouleurs's supersystem) but its great autonomy, without which it clearly cannot be treated as a new level in the structure of the Universe.

Some authors actually tried to investigate the properties of the Metagalaxy. One of these properties is the expansion of the Metagalaxy, as is evident from the red shift in the spectra of the galaxies. Rubin (1951) using the radial velocities of 70 galaxies and Ogorodnikov (1952) using those of 105 galaxies tried to determine the rotation parametes of the Metagalaxy

(the coordinates of the metagalactic pole). The results of the two authors are contradictory, which is either the result of insufficient material or is due to the fact that the observed galaxies are not members of a single rotating system.

The available data on the structure of the observable part of the Universe lead to the following two alternatives for the infinite expanses of the Universe: a) the Universe is made up of clusters of galaxies on various scales, which are distributed fairly uniformly, with natural fluctuations; b) the Universe is inherently hierarchic: each (relatively autonomous) system of n-th order is made up of a great number of systems of $(n-1)$-th order, and n is unlimited.

Alternative a gave rise to definite objections back in 19th century, which are known as Olbers's and Seeliger's paradoxes.

Olbers's paradox, or the photometric paradox, maintains that if we assume a uniform distribution of any kind of objects in an infinite Universe (allowing any distribution density), the surface brightness of the sky will not be less than the surface brightness of these objects. Indeed, let ϱ be the density of the particular type of objects in space, s the cross sectional area of these objects. The total solid angle occupied by all the objects then gives a divergent integral:

$$\int_0^\infty \frac{1}{r^2}\varrho \cdot 4\pi r^2\, dr = 4\pi s\varrho \int_0^\infty dr = \infty. \tag{83}$$

This shows that the objects inevitably overlap (an infinity of times), and therefore the resulting surface brightness of the sky should be equal to the surface brightness of these objects.

If we now assume that the uniformly distributed objects contain stars and pick out a single star in each object, we see that the stars are also uniformly distributed in the infinite space and the surface brightness of the sky should therefore be at least the surface brightness of stars, clearly a fantastic conclusion.

Seeliger's paradox, or the gravitational paradox, is essentially the same, although it refers to the gravitational forces applied to some point mass by the uniformly distributed objects. The resultant attraction vector is the infinite sum

$$\frac{Gm_1}{R_1^2}k_1 + \frac{Gm_2}{R_2^2}k_2 + \ldots + \frac{Gm_i}{R_i^2}k_i + \ldots, \tag{84}$$

where k_i is the i-th unit vector.

The main question now is whether series (84) converges or diverges. If the series is divergent, the applied force is infinite and we should observe infinite accelerations, which is of course not true.

The photometric and the gravitational paradoxes were discussed by numerous authors. Currently they can be resolved without much difficulty.

The gravitational paradox is resolved if we remember that the sum in (84) is a geometrical sum, so that the summation of the series does not lead to an expression of the form (83), which is the result of arithmetic summation. If a certain class of objects are uniformly distributed in infinite space, with only natural fluctuations, the resultant attraction due

to these objects is a random variable dependent on the natural fluctuations. The corresponding distribution function was derived by Holtsmark* (who incidentally solved the problem for electric fields, which in our context are entirely equivalent to gravitational fields):

$$\varphi(F)=\frac{2}{\pi F}\int_0^\infty e^{-\frac{\alpha x^{3/2}}{F^{3/2}}} x\sin x\,dx, \tag{85}$$

where $\alpha=(4/15)(2\pi GM)^{3/2}\cdot\varrho$, m is the mass of one object, ϱ is the average number of objects in unit volume.

It can be shown that for $F\to\infty$, Eq. (85) behaves as the function

$$\varphi(F)=\frac{15}{8}\sqrt{\frac{2}{\pi}}\,\alpha F^{-\frac{5}{2}}. \tag{86}$$

The probability of extremely large forces is thus zero. This completely resolves the gravitational paradox.

The photometric paradox is resolved using cosmogonic arguments. If we remember that the density of matter is finite everywhere in the Universe, no radiation process will go on indefinitely (whatever the actual radiation mechanism).

Therefore, assuming that the radiation reaching the Earth is made up of contributions from all the volume elements of the infinite space, we in fact maintain that a volume element distant T light years from the Earth was in a radiating state T years ago. This conclusion, however, is inadmissible, seeing that space is infinite.

Hypothesis b of universal hierarchy was developed by Charlier (1908) in order to avoid the photometric and the gravitational paradoxes.

As we have seen, these paradoxes can also be resolved assuming an infinite Universe uniformly populated with objects of any order.

The significance of the hierarchy hypothesis is therefore not that it resolves the photometric and gravitational paradoxes, but that this hypothesis is based on the cumulative experience of science: observations indicate that any assembly of objects combine to give a higher-order object with new qualitative properties, which apart from the objects under consideration also includes objects of other types as its constituent elements.

REFERENCES TO CHAPTER XXII

1. AGEKYAN, T.A. Theory of fluctuations in the number of observed galaxies.— Astron. Zh., 34:371. 1957.
2. AMBARTSUMYAN, V.A. Fluctuations in the number of extragalactic nebulae and galactic absorption.— Bull. Abastumani Astrophys. Observ., No.4. 1940.
3. AMBARTSUMYAN, V.A. On the nature of radio sources.— Trudy V Soveshchaniya po Voprosam Kosmogonii, 413. 1956.
4. AMBARTSUMYAN, V.A. Multiple galaxies.— Izvestiya AN Armenian SSR, 9:23. 1956.

* Holtsmark, G.J.— Ann. d. Physik 58:577. 1919.

5. AMBARTSUMYAN, V.A. The evolution of galaxies.— Izvestiya AN Armenian SSR, 11, No.5:9. 1958.
6. BAADE, W. The resoltuion of Messier 32, NGC 205, and the central region of the Andromeda Nebula.— Astron. Zh., 22:165. 1945; NGC 147 and NGC 185 — two new members of the Local Group.— Astron. Zh., 22:175. 1945. [Russian translations.]
7. VASHAKIDZE, M.A. A study of interstellar absorption from color excess of extragalactic nebulae, long-period cepheids, and other methods.— Bull. Abastumani Astrophys. Observ., No.13. 1953.
8. VORONTSOV-VEL'YAMINOV, B.A. Radio galaxies and galaxies with wide emissions in their spectrum.— Astron. Zh., 35:208. 1958.
9. VORONTSOV-VEL'YAMINOV, B.A. Interaction of galaxies and the nature of arms, bars, and tails.— Astron. Zh., 35:858. 1958.
10. VORONTSOV-VEL'YAMINOV, B.A. Atlas i katalog vzaimodeistvuyushchikh galaktik (Atlas and Catalogue of Interacting Galaxies). Part I. 1959.
11. KAVRAISKAYA, K.V. Determination of the distribution function of true flattening of galaxies.— Vestnik Leningrad University, No.1, Issue 1, p.146. 1958.
12. OGORODNIKOV, K.F. On the kinematics of the Metagalaxy.— Voprosy Kosmogonii, 1:150. 1952.
13. SHKLOVSKII, I.S. The radio emission of the Metagalaxy.— Trudy V Soveshchaniya po Voprosam Kosmogonii, 554. 1955.
14. SHAPLEY, H. Galaxies. — Harvard Univ. Press. 1961.
15. EIGENSON, M.S. Vnegalakticheskaya astronomiya (Extragalactic Astronomy).— Fizmatgiz. 1960.
16. ARP, H.C. Novae in the Andromeda nebula.— Astron. J., 61:15. 1956.
17. BAADE, W. and R. MINKOWSKI. Identification of the radio sources in Cassiopeia, Cygnus A, and Puppis.— ApJ, 119:206. 1954.
18. BAADE, W. and R. MINKOWSKI. On the identification of radio sources.— ApJ, 119:215. 1954.
19. BABCOCK, H.W. The rotation of the Andromeda Nebula.— Lick Obs. Bull., 19:41. 1939.
20. HOLMBERG, E. A study of double and multiple galaxies together with inquiries into some metagalactic problems with an appendix containing a catalogue of 827 double and multiple galaxies.— Lund. Ann., 6. 1937.
21. HOLMBERG, E. A photometric study of nearby galaxies.— Lund. Medd., No.128. 1950.
22. HUBBLE, E. Extra-Galactic nebulae.— ApJ, 64:321. 1926.
23. HUBBLE, E. The distribution of the extra-galactic nebulae.— ApJ, 79:8. 1934.
24. HUBBLE, E. The Realm of the Nebulae.— Oxford. 1936.
25. HUBBLE, E. The law of red-shift.— MN, 113:658. 1953.
26. HUMASON, M.L., N.U. MAYALL and A.K. SANDAGE. Red shift and magnitudes of extragalactic nebulae.— Astron. J., 61:97. 1956.
27. KERR, F.G., G.V. HINDMAN and B.G. ROBINSON. Observation of the 21 cm line from the Magellanic Clouds.— Austral. J. Phys., 7:297. 1954.
28. KRAUS, G.D. A preliminary study of the magnitude distribution of celestial radio sources.— Astron. J., 60:398. 1955.
29. MAYALL, N.U. Inclinations of spectrum lines in spirals.— Sky and Telescope, 8:3. 1948.
30. MAYALL, N.U. and L.H. ALLER. Rotation of Messier 33.— ApJ, 95:5. 1942.
31. MILLS, B.Y. Radio frequency radiation from external galaxies.— Handbuch der Physik, 53:239. 1959.
32. MINKOWSKI, R. The problem of the identification of extragalactic radio sources.— Publs. Astron. Soc. Pacif., 70:143. 1958.
33. MORGAN, W.W. and N.U. MAYALL. A spectral classification of galaxies.— Publ. Astron. Soc. Pacif., 69:291. 1957.
34. MORGAN, W.W. A preliminary classification of the forms of galaxies according to their stellar population.— Publ. Astron. Soc. Pacif., 70:364. 1958.
35. NEYMAN, G. and E.L. SCOTT. A theory of the spatial distribution of galaxies.— ApJ, 116:144. 1952.
36. NEYMAN, G., E.L. SCOTT and C.D. SHANE. On the spatial distribution of galaxies: A specific model.— ApJ, 117:92. 1953.
37. PAGE, T. Radial velocities and masses of double galaxies.— ApJ, 116:63. 1952.
38. PAWSEY, G.L. Sidney investigations and very distant radio sources.— Publ. Astron. Soc. Pacif., 70:133. 1958.
39. RUBIN, V.C. Differential rotation of inner metagalaxy.— Astron. J., 56:47. 1951.
40. SANDAGE, A. Current problems in the extragalactic distance scale.— ApJ, 127:513. 1958.
41. SHANE, C.D. and C.A. WIRTANEN. The distribution of extragalactic nebulae.— Astron. J., 59:285. 1954.
42. SHANE, C.D. The distribution of extragalactic nebulae.— Astron. J., 61:292. 1956.
43. SHAPLEY, H. Seventy six thousand faint galaxies.— Harvard Bull., 889, 1932.

44. SHAPLEY, H. and A. AMES. A survey of the external galaxies brighter than thirteenth magnitude.— Annals Harv. College, 88, No. 2, 43. 1932.
45. SHAPLEY, H. Thirty-six thousand galaxies in the south polar cap.— Harvard Annals, 105:137. 1937.
46. SHAPLEY, H. The Inner Metagalaxy. 1957.
47. SCHWARZSCHILD, M. Mass distribution and mass-luminosity ratio in galaxies.— Astron. J., 59:273. 1954.
48. SPITZER, L. and W. BAADE. Stellar populations and collisions of galaxies.— ApJ, 113:413. 1951.
49. STEBBINS, J. and A. E. WHITFORD. Magnitudes and color of 176 extragalactic nebulae.— Apj ApJ, 115:284. 1952.
50. VAUCOULEURS, G., de. Evidence for a local supergalaxy.— Astron. J., 58:30. 1953.
51. VAUCOULEURS, G., de. Studies of the Magellanic Clouds I. Dimension and structure of the Large Cloud from star counts and long-exposure photographs.— Astron. J., 60:126. 1955.
52. VAUCOULEURS, G., de. Studies of the Magellanic Clouds II. Dimensions and structure of the Small Cloud from star counts and long-exposure photographs.— Astron. J., 60:219. 1955.
53. VAUCOULEURS, G., de. Studies of the Magellanic Clouds III. Surface brightness, colors and integrated magnitudes of the Clouds.— Astron. J., 62:69. 1957.
54. VAUCOULEURS, G., de. The distribution of bright galaxies and the local supergalaxy.— Vistas in Astronomy, London—New York, 1584, 1956.
55. VAUCOULEURS, G., de. Survey of bright galaxies south of −35° declination.— MN, 117:225. 1957.
56. VAUCOULEURS, G., de. Ring structure in galaxies as distance indicators.— Astron. J., 62:13. 1957.
57. WYSE, A. B. and N. U. MAYALL. Distribution of mass in the spiral nebulae Messier 31 and Messier 33.— ApJ, 95:24. 1942.
58. WYATT, S. P. and F. C. BROWN. Position angles and shapes of galaxies in Cetus.— Astron. J., 60:415. 1955.
59. ZWICKY, F. Statistics of clusters of galaxies.— Proc. of the Third Berkeley Symposium on Mathematical Statistics and Probability, 3:113. 1956.
60. ZWICKY, F. Multiple galaxies.— Ergebnisse der exakten Naturwiss., 29:344. 1956.
61. ZWICKY, F. Morphological Astronomy.— Berlin — Gottingen — Heidelberg. 1957.

NAME INDEX

ABBE 671
ABELL 692
ABT 229, 237
ADAMS 31, 32, 49, 53, 156, 160, 162, 167—170, 218
AGEKYAN 83, 480, 496, 522, 600, 667, 683, 703
AIRY 512
AITKEN 58, 59, 69, 83
ALBADA 210
ALEXANDER 174
ALLER 218, 297, 306, 330, 412, 413, 414, 415, 423, 665, 704
AMBARTSUMYAN 50, 79—81, 83, 146, 208, 238, 244, 246, 247, 248, 250, 268, 282, 284, 291, 297, 326, 328, 330, 340, 352, 356, 361, 378, 391, 400, 409, 412, 413, 415, 418, 423, 437—442, 478—481, 496, 522, 547, 552, 590, 593, 597, 600, 612, 613, 614, 615, 630, 680, 684, 685, 703, 704
AMES 678
ANROTO 185
ARAKELYAN 250
ARGELANDER 139
ARP 200, 205, 214, 215, 666, 677, 704

BAADE 188, 189, 199—202, 211, 214, 218, 288, 305, 489, 497, 624, 630, 645, 661, 662, 676, 677, 690, 699, 705
BABCOCK 208, 237, 250, 255, 665, 704
BADALYAN 159, 202, 205, 218
BAILEY 155, 174, 196
BAKER 330, 405, 406, 423
BALASZ 174
BALDWIN 678, 692
BALLARIO 612
BARBIER 31, 232, 237, 420, 424, 434, 442
BARKHATOVA 607, 630
BARNARD 320
BARTAYA, 53
BASBRIDGE 386, 391
BAUM 663
BEALS 36, 425, 427, 435, 442
BECKER 155, 176, 194, 208, 497
BELOPOL'SKII 28, 58, 139, 160, 168, 184, 235, 254, 272
BERDICHEVSKAYA 356
BERGH, Van den 189
BERMAN 250, 301
BERNHEIMER 215
BESSEL 57, 71
BIDELMAN 42
BINNENDIJK 510
BLAAUW 176, 177, 178, 186, 191, 613, 614
BLAGG 127
BLAZHKO 58, 96
BLEKSLEY 152
BOEV 76
BOK, B. 327, 467, 497, 626
BOK, P. 497
BOLTON 692, 696
BOS, Van den 59
BOTTLINGER 47, 208
BOWEN 392, 423
BOYARCHUK 545
BRILL 13, 467, 497
BROWN 678, 692, 696, 700, 705
BURBIDGE 442, 624
BURNHAM 58

CAMM 198, 200, 547
CAMPBELL 147, 547
CANAVAGGIA 155, 160, 218
CANNON 160
CEDERBLAD 479, 497
CHALONGE 31, 54, 232, 434, 442
CHANDRASEKHAR 155, 338, 351, 356, 361, 378, 391, 427, 442, 480, 545, 548, 555, 577, 600, 626
CHARLIER 610, 703
CHERNOVA 491
CILLIE 405, 406, 423
CLARK 558
CODE 157, 492
COUTEAU 82, 83
CRAMER 467
CURTIS 320, 631

DADAEV 123, 128, 132, 138
DAMBARA 153
DARTAYET 215
DARWIN 118
DEICH 47, 83, 187, 547
DEMBOWSKI 58
DESSY 215
DETRE 174, 175
DEURINCK 169
DOBROVOL'SKII 325
DOMBROVSKII 290, 314, 423, 424
DUGAN 117, 132

EDDINGTON 49, 125, 193, 207, 356, 374, 375, 391, 520, 548
EFREMOV 146, 651
EGGEN 16, 83, 132, 149, 155, 159, 160, 177, 201, 202, 216, 614
EIGENSON 704

FEAST 674
FEDOROVICH 159
FESENKOV 497
FETLAAR 105—108, 138
FINSEN 69
FLORYA 138, 146, 151, 175, 473, 497, 624, 630
FOWLER 389, 391
FRANK-KAMENETSKII 210, 218
FRANKLIN 476
FRICKE 666
FROST 184

GALKIN 53
GAMALEI 625
GAPOSCHKIN, S. 51, 53, 147, 174, 182, 194, 218
GAZE 303, 330, 369, 391, 392, 423, 492
GERASIMOVICH 427, 442
GERNER 625
GITS 150
GLAZENAP 58, 63
GLIESE 463
GOLDBERG 330, 407, 423
GOODRICKE 56, 93
GORBATSKII 237, 250, 253, 454, 435, 442
GORDELADZE 284, 328, 478, 497
GORDON 182, 210, 218
GOULD 496
GREENSTEIN 276, 420, 424
GRIGOR'EVA 188
GRIGORYAN 231, 237
GUK 414, 423
GÜNTZEL-LINGNER 69, 83
GUREVICH 158, 218, 285, 584, 590, 594, 598, 600, 630
GURZADYAN 300, 418, 424
GUSSOV 193
GUTHNICK 149, 184

HALL, R.G. 453
HALL 314
HARDY 176
HARO 248, 249, 250, 283, 649
HAZARD 678, 692, 696, 700
HEBB 330, 412, 423
HENYEY 383, 391
HERBIG 201, 250, 256, 283
HERSCHEL, J. 57, 58, 672, 673
HERSCHEL, W. 56—58, 78, 443, 601, 605, 683
HERTZSPRUNG 27, 29, 171, 196, 212, 214, 445, 451
HERZOG 680, 693
HILL 495
HILTNER 314, 323, 613, 677
HINDMAN 493, 674, 704
HIPPARCHUS 443
HOENBERG 202
HOGG 608
HOLMBERG 641, 645, 648, 666, 683, 695, 704
HOLTSMARK 391, 703
HOOF 169, 184
HOPF 338, 356
HOUTGAST 387, 391
HUANG SU-SHU 52, 54, 127, 138, 173, 544, 548
HUBBLE 189, 275, 303, 320, 497, 631, 632, 637, 640, 641, 644, 646, 651, 653, 659, 660, 663, 666, 678, 704
HUFFER 127, 132, 323
HUGHES 168
HULST 205, 317, 492, 497, 538, 665, 677
HUMASON 53, 250, 514, 647, 649, 656, 658, 704
HUSSEY 58

IKAUNIEKS 490, 497
INNES 66, 74, 79
IRANOVA 250
IROSHNIKOV 210
IRWIN 85, 90, 92, 666
IWANOWSKA 176, 177, 218

JACKSON 58, 83
JACOBSEN 167, 168, 183
JEANS 553, 555, 600
JENKINS 26
JOHNSON 16, 35—36, 41, 42, 54, 449, 451, 618
JOY 31, 49, 53, 156, 160, 182, 188, 191, 211, 213, 214, 218, 237, 250
JOYNER 9, 10, 14, 17, 458, 498

KALANADZE 53
KALINYAK 487, 492, 497
KAMP, Van de 50, 53, 250
KAPTEYN 75, 78, 454, 457, 464, 515, 516, 517, 519
KAVRAISKAYA 653, 704
KEENAN 8, 36, 40, 41, 53, 54, 157, 237
KELLMAN 41, 42, 54, 157
KERR 493, 498, 674, 704
KHABIBULLIN 497
KHARADZE 497
KHOLOPOV 146, 189, 200, 202, 250, 497, 615, 621, 630
KING 85, 91, 92
KINMAN 624, 625
KIPPER 165, 201, 210, 330, 421, 424
KO 692, 701
KOELBLOED 673
KOHLSCHUTTER 31
KOLMOGOROV 306
KOPAL 51, 104—108, 132—133, 138, 195
KOPYLOV 53, 545
KOVAL'SKII 63, 512

KOZYREV 155, 207, 268, 281
KRAFT 194, 204, 218, 701
KRASOVSKII 487, 492, 497
KRAT 105, 129, 132, 138
KRAUS 692, 700
KRON 16, 132, 160, 216, 218
KRÜGER 210
KUIPER 6, 21, 23, 24, 51, 81, 171, 195, 463, 497
KUKARKIN 138, 146, 147, 151—153, 157, 168, 173, 177, 188, 189, 192, 198—200, 211, 218, 227, 237, 241, 250, 258, 489, 490, 497, 624, 630
KULIKOVSKII 69, 83, 151
KUNITSKII 33, 497
KUZMIN 489, 497, 554, 558, 565, 600
KWEE 538, 543, 548

LAMMA 22
LEAVITT 196, 671
LEBEDINSKII 158, 218, 284, 285, 614
LEHMAN 168
LEVIN 590, 594, 600, 630
LINDBLAD 35, 527, 558, 564, 600, 666
LINDGRAN 623
LUDDENDORF 214, 237
LUNDMARK 47, 54, 659, 683
LUST-KULKA 181
LUYTEN, W. 80, 83, 454, 464

MATT 701
MARKARYAN 480, 481, 497, 606, 612, 614, 630, 680
MARTYNOV, 21, 134, 138, 146
MASEVICH 49, 54, 195, 453
MAUNSELL 30
MAURY 33, 127
MAYALL 237, 514, 623, 624, 647, 649, 650, 656, 658, 663, 664, 665, 667, 668, 704, 705
MAYER 56
McLAUGHLIN 237, 260
McNAMARA 185, 186, 192
MEINEL 35, 54
MEL'NIKOV 20, 27, 30, 35, 51, 53, 154—156, 159, 167, 169, 170, 191—195, 198—199, 202, 208, 214, 218, 312, 365, 366, 391
MELOTTE 601
MENDOZA 605
MENZEL 259, 297, 330, 405, 406, 412, 414, 423
MERRILL, J. 104, 138, 250
MERRILL, P. 221, 222, 237
MICHARD 21
MICHELL 56
MITCHELL 604
MICZAIKA 30
MILFORD 21, 22
MILLS 495, 675, 697
MILNE 125, 389, 391, 531
MINEUR 198, 200, 626, 628
MININ 418, 424, 437
MINKOWSKI 290, 293, 422, 424, 697, 699, 704
MLODZEEVSKII 63, 69
MOORE 48
MORGAN 8, 16, 24, 35—36, 40, 54, 157, 176, 191, 446, 448, 449, 451, 492, 632, 650, 663, 704
MULLER 492, 494, 538, 543
MÜNCH 176, 351, 356, 480, 545, 548
MUSTEL 23, 270, 271, 291, 347, 348, 353, 356, 442

NEYMAN 680, 683, 704
NIKITIN 407, 423
NIKONOV 18, 40, 487, 492, 497
NORDSTROM 515

OGORODNIKOV 168, 218, 325, 477, 497, 531, 533, 547, 587, 597, 600, 654, 701, 704
OHLSSON 444
OKE 36
OLBERS 680
OLDEN 46
OORT 205, 237, 487, 492, 498, 526, 531, 538, 544, 548, 555, 563, 565, 600, 621, 646, 680
OOSTERHOFF 212
OSKANYAN 250
OSTERBROCK 492

PAGE 420, 424, 687, 704
PANNEKOEK 167, 170, 210, 375, 390, 391, 673
PARENAGO 31, 49—51, 54, 138, 146, 147, 150, 155, 158, 159, 173, 188, 190, 192, 194, 201, 202, 218, 241, 250, 258, 324, 446, 451, 453, 473, 489—492, 497, 514, 517, 521, 522. 535, 547, 564, 565, 600, 619, 624, 680
PARIISKII [PARIJSKY] 168
PAVLOVSKAYA 191, 202, 491, 514
PAYNE-GAPOSCHKIN 147, 157, 158. 164, 182, 184, 193, 212, 213, 217, 218, 275, 291
PEARCE 515
PECKER 155, 218
PELS-KLUYVER 180, 181
PETRIE 27, 186
PETROV 138
PICKERING 139
PIGOTT 139
PIFAGETTA 671
PIKE 125, 129
PIKEL'NER 317, 330, 495
PIOTROVSKI 105, 106, 138
PLASKETT 250, 515
PLUMMER 620, 630
POPPER 23—24, 47, 132
PRAGER 142, 147
PRESTON 177, 218
PROKOF'EVA 154
PSKOVSKII 202, 218

RABE 69, 83
RAIMOND 665, 677

RAYLEIGH 327
REDDISH 200, 201, 218
REDMAN 646
RHIJN 454, 458, 459, 463, 464, 498
RICCIOLI 56, 78
RICHMOND 458
RITCHEY 631
ROBERTS 195
ROCHE 119
ROGERS 201
ROMAN 42, 612
ROSSELAND 218
ROSSITER 59
ROUGOOR 544
RUBIN 701, 704
RUDNITSKI 497, 547
RUSAKOV 480, 497
RUSSELL 27, 48, 54, 85, 91, 92, 96, 97, 99, 104, 108, 120, 126, 138, 367, 391, 445

SAFRONOV 489, 497, 565, 600
SANDAGE 195, 618, 621, 624, 630, 647, 649, 656, 658, 662, 663, 704, 705
SANFORD 178, 182
SAVEDOV 180, 186
SAWYER 189, 210, 212
SCHIAPARELLI 58, 456
SCHLESINGER 85, 92
SCHNELLER 143, 147
SCHUSTER 356, 357
SCHWARZSCHILD 85, 90, 170—172, 210, 213, 466, 467, 584, 693
SCHWARZSCHILD, B. 162, 168, 170
SCHWARZSCHILD, K. 356, 357, 520, 548
SCHWARZSCHILD, M. 92, 162, 167, 168, 170, 209, 218, 522, 694, 705
SCOTT, E. 680, 683, 704
SCOTT, R. 226, 237
SEARES 9, 10, 14, 17, 458, 459, 467, 498
SEATON 409, 412, 423, 424
SEELIGER 456, 457
SEMENOVA 104, 138
SENTSOVA 150
SEVERNYI 194, 207, 218, 442
SEYFERT 649
SHAIN, C.A. 495
SHAIN, G.A. 54, 83, 224, 237, 303, 307, 309, 369, 372, 391, 423, 440, 442, 544
SHAIN, P.F. 391
SHAKHBAZYAN 624
SHAKESHAFT 692
SHANE 680, 683
SHAPLEY 160, 196, 197, 199, 200, 211, 487, 617, 623, 630
SHARBE 105
SHARPLESS 492
SHATSOVA 454, 497
SHCHEGOLEV 646
SHIH SHENG 443
SHKLOVSKII 290, 299, 302, 317, 423, 424, 492, 495, 692, 699, 704
SHMIDT 18
SHNIREL'MAN 491
SHORTLEY 330
SHTERNBERG 58
SITTERLY 105, 138
SLEE 495, 696
SLETTEBAK 544
SLIPHER 663
SLYUSAREV 352, 356, 438, 442
SMITH, D. 16
SMITH, H. 211
SOBOLEV 224, 237, 250, 253, 256, 259, 269, 297, 391, 423, 424, 442, 453, 547
SOFRONITSKII 122
SONN 497, 547
SPITZER 421, 424, 522, 584, 600, 690, 705
STANLEY 696
STAHR-CARPENTER 493
STEBBINS 16, 17, 127, 153, 154, 155, 162, 179, 448, 487, 648, 705
STERNE 151, 175
STIBBS 9, 15, 54, 160
STOYANOVA 35
STRAND 46, 453
STROMBERG 528
STRÖMGREN, 312, 315, 316, 391
STRUVE, L. 58
STRUVE, O. 51, 52, 54, 78, 127, 128, 138, 157, 173, 176—177, 184, 190, 208—209, 218, 240, 250, 252, 253, 256, 302, 372, 391, 544
STRUVE, W. 57, 59, 320, 457
SVOLOPOULOS 160, 204, 216, 218
SWEENY 676
SWINGS 240, 250, 302
SYCHEVA 154

TAYLOR 615
TERRAZAS 176
THACKERAY 190, 674
THIESSEN 167, 194, 195
THORNTON 687
TIKHOV 516
TORGOR 205
TRUMPLER 47, 48, 320, 498, 548, 605, 607, 608, 630
TSESEVICH 102, 103, 105, 108, 109, 138, 146, 175
TSOI DYAI 232, 237
TYCHO BRAHE 285

UNDERHILL 24, 351, 356

VASHAKIDZE 290, 497, 680, 704
VAUCOULEURS 637, 645, 666, 672, 674, 705

VORONTSOV-VEL'YAMINOV 147, 256, 259, 274, 276, 285, 291, 298, 330, 392, 414, 423, 438, 442, 447, 491, 492, 497, 685, 686, 692, 700, 704

WALDMEIER 11, 54
WALKER 273, 604
WALLENQUIST 79, 83, 623
WALRAVEN 168, 170, 175, 178, 181, 210, 218
WALTON 160
WEAVER 33, 492, 498, 548
WEENEN 437, 442
WEMPE 21
WERNER 505
WESSELINK 190, 211, 218
WESTERHOUT 493, 494, 498, 543, 548
WESTFOLD 692
WESTGATE 544
WEYL 149
WHITFORD 16, 323, 448, 487, 492, 648, 705
WIERZBINSKI 21, 83
WILD 680, 693
WILLIAMS 185, 680
WILSON 160, 237
WIRTANEN 680, 683, 704
WOERDEN, Van 665, 677
WOLF 475, 476
WOLTJER 118
WOOLLEY 9, 15, 614
WURM 299
WYSE 108, 138, 668, 705

ZAITSEVA 160
ZANSTRA 296, 300, 398, 399, 417, 418, 423, 424, 437, 442
ZEIPEL 620, 623
ZHEVAKIN 208, 218, 226, 237
ZHURAVLEV 365, 391
ZURHELLEN 85, 91, 92
ZVEREV 138, 146
ZWICKY 644, 652, 680, 683—685, 687, 689, 693, 695, 705

SUBJECT INDEX

Absolute brightness 1
  magnitude effect 390
  stellar magnitude 1, 25ff, 445
    motion 502
Absorption coefficient 468, 469
  function 469
Antapex 507
Apex 506, 507, 511
Apparent density function 472
  sphericity 653
  stellar magnitude 1
Associations, stellar 247, 441, 612
Asymmetry of stellar motions 525
Atmospheres, cepheids 170ff
  RV Tau stars 229
  stellar 253, 332, 357ff, 425ff
    chemical composition 366

Binaries 27, 44, 51, 55ff, 248 249
  evolution 133ff
  masses 69
  models 94ff, 118, 119, 122
  photometric 55, 93ff
  spectroscopic 55, 84ff, 425
  visual (astrometric) 55, 56ff
Binary encounters 577ff
Bolometric correction 7
  luminosity 3
Boltzmann equation 553
Bright nebulae 293
Brightness, absolute 1
  distribution over the disk 339
Broadening of spectral lines 369

Carbon stars 245
Cataclysmic stars 145
Catalogue of variable stars 142
Centroid 518
Cepheids 51, 144, 659
  distribution over the sky 188
  dwarf 211
  extragalactic 210, 217
  frequency of periods 189
  in globular clusters 210
  in Magellanic Clouds 210, 215
  long-period 144, 148, 188ff

Cepheids, long-period, Population I 177, 182, 188
    Population II 177, 182, 188
  motion 190
  period—density relation 192
    — luminosity relation 193
    — spectrum relation 191
  Population I stars 188, 214
  Population II stars 188, 214
  pulsation theory 205
  short-period (also see RR Lyrae stars) 144, 173, 188ff, 619
    Population I 177, 188
    Population II 177, 188
  spatial density 188
Circular orbits of stars 566
Classification of binaries 55
  of close pairs 129ff
  of clusters of galaxies 687
      Herzog, Wilde and Zwicky's catalogue 693
  of eclipsing binaries 51
  of galaxies, Hubble 631
    Morgan 650
    Vaucouleurs 637
  of globular clusters, Shapley's 617
  of long-period variables 219
  of nebulae 295
  of novae 257
  of nova light curves 260
  of open clusters, Markaryan's 606
    Trumpler's 604
  of planetary nebulae 295
  of spectra of typical novae 264
  of variables 143
  spectral 32ff, 39ff, 387
Clouds of clusters of galaxies 687
  of galaxies 687
Clusters, globular 196, 549
  of galaxies 687, 693
    compact 693
    in Coma Berenices 688ff, 695
    in Virgo 690
    masses 693
    open 693
    radio sources 700
    regular 687
    semicompact 693

Clusters, open 79, 549
Coal sacks 325, 327, 477
Coherent scattering 359
Collision of galaxies 319, 699
Color excess 20, 470
— luminosity diagram 448
Colorimetric systems 16
Colors, stellar 14ff
Coma Berenices cluster 688ff, 695
Compact clusters of galaxies 693
Contact systems 51
Continuous emission 247
Coordinates, galactic 443
Crab Nebula 288, 290, 319, 422
Critical velocity of a star, mean 574
Cumulative errors 151

Dark companions 71
Differential line shifts in cepheid spectra 167
rotation of a star system 531
Diffuse interstellar matter 292ff
nebulae 247, 302ff, 392, 414, 479
Dilution coefficient 393
Disk component of the Galaxy 188
Dissipation of stellar systems 590
Distance modulus 445
Distances, extragalactic 645, 659
Distribution, ellipsoidal 520
Maxwellian 519
of flattening of galaxies 652
of galactic types 651
of galaxies 678
of residual velocities 567, 585, 597
Doppler effect 655
Double galaxies 683
masses 686
stars, see Binaries
Dual periodicity in variables 141
Dust nebulae 292, 320ff
bright 328
dark 325
Dwarf cepheids 211
Dwarfs 248, 283, 366, 371, 390, 441
red 447
white 243, 447
Dynamic definition of a stellar system 596
orbital elements in a binary 63
parallaxes of binaries 27, 70

Eclipsing binaries, see Photometric binaries
classification 51
Effective temperature 342
wavelength 8
Ejections from stars 127, 169, 234, 266, 268, 299, 314, 330, 435, 441
Ellipsoidal velocity distribution 520
Elliptical galaxies 631, 633, 637
Emden equation 622
Emission bands 425
lines, intensity 430
profiles 427ff
Envelopes of long-period variables 225
of stars, ejection 266, 268
Equator, galactic 443, 488
Equilibrium of stellar photospheres 332
Evolution of stars 133ff
of stellar systems 594, 596
Evolutionary sequence, of rotating systems 596
Expansion of space 655, 701
Extended stellar envelopes 254
Extragalactic distances 645, 659
Baade scale 661, 662
Hubble scale 660
Sandage scale 662

Filamental nebulae 304, 314
Fission of galaxies 700
Flares, stellar 146, 270, 284
Fluctuation theory 481, 482
Fluorescence of stellar envelopes 224
Forbidden lines in nebular spectra 407
Fundamental equation of stellar dynamics 552, 555, 560
plane 576

Galactic arms 205, 317
center 487, 536
circle 443
clusters 601
component, disk 188, 490
halo 188, 490, 491
intermediate 489, 490
concentration 443
coordinates 443
density 443
diameter 646
equator 443, 488
latitude 444
longitude 444
nebulae 293
nucleus 631
mass 667
rotation 663
spiral galaxies 649
plane 443
poles 444
subsystems 489
Galaxies 631ff
clouds 687
of clusters 687
clusters, see Clusters of galaxies
collision 319, 699
distribution 678
groups 687, 691
irregular 631, 637

Galaxies, masses 667
lenticular 637
multiple 683
normal 697
regular 631
ring-shaped 637
spiral 317
Galaxy 443
boundary 488
center 487, 536
radio observations 492
rotation 529, 538
spiral structure 491
structure 486
subsystems 489
total number of stars 459
Gaseous envelope of a star 267
nebulae 293ff, 392
bright 302
ring around stars 252
Giants 143, 283, 390, 447
Globular clusters 189, 196, 214, 486, 549, 551, 601, 615ff
Globules 292, 327
Gravitational acceleration on star surface 150, 359
field of a stellar system 549
paradox 702
Group parallaxes of stars 27, 612

Halo subsystem 188
Hertzsprung—Russell diagram 447
Hierarchic structure of the Universe, hypothesis 703
H I and H II regions 315
Holmberg's criterion for multiple systems 693
Hubble constant 640, 655, 658, 661
Hubble's law 655
Hyades 612
Hydrodynamic equations for stellar systems 558
Hydrogen, interstellar 317, 492, 538

Incoherent scattering of light 383
Integral apparent magnitude function 455
Integrated luminosity 3
Intermediate spirals 632
Interstellar absorption 18, 310, 320, 468
dust 320ff
gas 303, 310, 317, 492, 538
Ionization in nebulae 400
Irregular field of a system 549
forces 549, 575
galaxies 631
variables 143, 231
Isophotic wavelength 8

K effect 514
Kapteyn's selected areas 515
star streams 519, 520
Kirchhoff—Planck law 343

Latitude, galactic 444
Lenticulars 637
Light curve 93
cepheids 148, 173
novae 259
reflected 97
RR Lyrae stars 173
standard 149, 150
Limb darkening 103
Line intensities in RR Lyrae stars 180
intensity curve of long-period cepheids 168
of RR Lyrae stars 180
of nodes in a binary system 63
Liouville theorem 553
Local Group 496, 691
Supersystem 691
Longitude, galactic 444
Long-period cepheids 148, 188ff
variables 144, 219, 439
Population I 177, 182, 188
Population II 176, 182, 188
Luminosity, blackbody 4
classes 35
function 453
of galaxies 641, 644
integrated (bolometric) 3
nebula 392
stellar 143, 445

Magellanic Clouds 189, 190, 196, 210, 215, 275, 279, 305, 319, 671
Magnetic fields in stars 254
variables (α^2 CVn) 235
Main sequence 446, 522, 603
Masses of binaries 69
of double galaxies 686
galaxies 667
in clusters 693
nebulae 413ff
novae 272
stellar 44ff
Mass—luminosity diagram 451
Maxwell's distribution 519
Metagalaxy 655, 680, 701
expansion 655, 701
Metastable states 407
Method, Airy-Koval'skii (determination of Sun motion) 512
Irwin (determination of orbits of spectroscopic binaries) 90
Lehmann-Filheś (determination of orbits of spectroscopic binaries) 88
Mlodzeevskii (determination of orbits of binaries) 63
Oort—Vashakidze (star density determination) 474
Schwarzschild (determination of orbits of spectroscopic binaries) 90
Thiele—Innes (of orbits of visual binaries) 66

Monochromatic radiative equilibrium 359
Moving clusters 610
Multiple galaxies 683
stars 55ff, 77, 549
systems, Holmberg's criterion 683

Nebulae, dark 292, 475
diffuse 247, 302, 392, 478
masses 414
dust 292, 320ff
bright 292, 328
dark 325
galactic 293
gaseous 293ff, 392
planetary 293ff, 317, 392, 399
masses 413
radiation pressure 415
Nebular lines 392, 398
part of emission spectrum 239
spectrograph 303
stage of a typical nova 264
Nebulium 392, 399
Neutral hydrogen 538
Novae 145, 238, 257ff
in extragalactic nebulae 275ff
light curves 259, 260
masses 272
permanent 258
reasons for eruption 284
recurrent 257, 258
spectral stages 262
typical 257, 262
Nova-like stars 238, 257, 278
Non-stable stars 127
Nonstationary state of a star system 551
Nucleus of the Galaxy 492
of planetary nebulae 295, 299, 300, 392, 437

O-associations 612, 613
Olbers paradox 681, 702
Oort coefficients 534, 566
Open clusters 549, 601
moving 610
of galaxies 693
star clusters 79, 601
Optical depth 324
Optical thickness 469
Orbits, spectroscopic binaries 84ff
visual binaries 62

Parallax 25ff
average annual 510
secular 508
dynamic 70
group 612
spectroscopic 391
Peculiar spectrum 699
stellar motion 502
Period—luminosity curve 660
of a variable 141
Permanent novae 258
Phase effect 124
Photoelectric colorimetric systems 16
Photometric binaries 28, 44, 51, 93ff
paradox 681, 702
Photosphere, stellar 332
Planetary nebulae 276, 293, 294, 317, 392, 399
masses 413
nucleus 295, 299, 300, 392, 437
star-like 295
Pleiades 449, 603, 606, 612
Population I stars 188, 214, 450, 489, 603, 618, 674
Population II stars 188, 214, 489, 618, 697
Principal nebular lines 392
Proper motion of stars 517, 529
Pulsation theory 205, 226
Pulsations of cepheid 207

Quasistationary state of stellar system 551, 597

Radial velocities of stars 500, 517, 529
velocity curve 229
long-period cepheids 160
RR Lyrae stars 178
Radiant equilibrium of stellar photospheres 332
of a cluster 610
Radiation flux 2, 20, 445
pressure in nebulae 415
Radiative equilibrium, monochromatic 359
Radioastronomic methods 220, 290, 317, 318, 674, 692, 696
Radio brightness, relative 697
galaxies 696, 697, 698
magnitude of a source 696
nebulae 290, 291, 697
Radius, stellar 6
Recurrent novae 145, 257, 258
Red semi-regular variables 144, 230
shift 47, 655, 701
Reflected light curve 97
Regular field 549
forces 549
variables 143
Relaxation time 584, 598
Residual velocity distribution function 567, 585, 597
Ring-shaped galaxies 637
Rosseland theorem 396
Rotation, galaxies 663
Galaxy 529, 538
stars 371, 544
RR Lyrae stars 173, 188ff, 212
Population I 177, 188
Population II 177, 188

Schwarzschild—Schuster model of stellar envelopes 357, 378
Seeliger paradox 702
theorem 455
Selective absorption 469
coefficient 20, 469, 470
Semiregular variables 143, 227, 230
Sequence, globular clusters 618
main 446, 522, 603
white-blue 276, 447
Spatial velocity of stars 499, 505, 525
Spectral lines, broadening 369
sequence, stars 387
stages of novae 262
types of stars 32ff, 39ff, 445
Spectroscopic binaries 55, 84ff, 425
parallaxes 29, 391
variables 232, 251ff
Spectrum—luminosity diagram 445
nebular, continuous 419
—spin diagram 544, 545
typical novae 262
variation curve, long-period cepheids 180
RR Lyrae stars 176
Sphericity, apparent 653
Spiral arms, coiling hypothesis 666
galaxies 317, 631, 632
barred 633
galaxy in Andromeda 675
structure of the Galaxy 491
Stability of star clusters 625ff
Standard light curves of cepheids 149, 150
Star clusters 595ff, 610
globular 601, 615ff
moving 610
open 601
stability 625ff
density function 463, 472
motion, absolute 502
asymmetry 525
systems 549
Stark effect 369
Stars, Algol-type 93, 129, 130, 131
Be 145, 426, 433, 434
binary, see Binaries
double, see Binaries
late-type, with bright lines 439
novae, See Novae
nova-like, See Nova-like stars
P Cygni 279, 425, 433
rotation 371, 544
RR Lyrae (see also RR Lyrae stars) 173
semiregular, red 144
yellow 144
Trapezium-type 79
UV Ceti 249
Stars, variable, see Variables
with extended atmospheres 254
with variable bright lines 254
Wolf—Rayet 368, 425, 433, 435ff
Stationary state of star system 551, 596
Stellar associations 247, 441, 612
atmospheres 253, 332, 357ff, 425ff
chemical composition 366
diameter 6
dynamics 549ff
fundamental equation 551, 555, 560
evolution 133ff
magnitude, apparent 1
motion, absolute 502
peculiar 502
proper 502, 517, 529
rotational 544
photospheres 332, 357ff
equilibrium 332
Population I 188, 214, 450, 489, 603, 618, 674
Population II 188, 214, 489, 618, 697
statistics, first equation 465ff
second equation 466
Stokes operator 553
Stratification of radiation 404
Subdwarfs 131, 138, 447
Subgiants 44, 50, 51, 130, 137, 282, 447
Sun, position in the Galaxy 488, 491, 536
Supergiants 203, 281, 283, 351, 365, 371
Supernovae 1, 145, 257, 285ff
type I 286
type II 286
Supersystem of galaxies 691, 692
Surface acceleration curve, cepheids 165
long-period cepheids 180

T-associations 247, 441, 615
Temperatures, cepheids 171
nebulae 411
stellar 387ff, 396, 438, 439
effective 343
Thermodynamic equilibrium, local 341
Tidal deformation of galaxies 685
Trapeziums 79
Trigonometric parallax 25
Triple stars 77
True density function 472
Turbulence in cepheid atmospheres 171
In interstellar dust 326
in nebulae 306
Two-stream theory, Kapteyn's 519

Variable magnetic fields in stars 254
Variables 490
Algol-type 93
δ Cep, see Cepheids
β CMa 184

Variables, designation 142
eclipsing, see Photometric variables
irregular 143, 231
late-type 144
long-period 144, 219, 439
low-luminosity 143
magnetic, α^2 CVn 235
nova-like 238, 257, 278, 282
physical 139
pulsating 217
RCrB 146, 244
regular 143
RV Tau 144, 227
RW Aur 146, 246
semiregular 143, 227
late-type 144
red 230
yellow 230
spectroscopic 232, 251ff
T Tau 246, 441
U Gem 145, 241
UV Cet 146
W Vir 177, 182
Variables, Z And (nova-like) 145, 238, 439
Z Cam 145, 244
α^2 CVn 145, 235
Velocity, centroid 517, 518
distribution function 560
rotational, apparent 544
true 544
stellar, critical 574, 590
peculiar 519
radial 500, 517
residual 517, 518
spatial 499, 505, 525
Vertex 519
Virgo cluster 690
Virial theorem 572
Visual binaries 56ff
orbits 62
wide pairs 75

White dwarfs 243
-blue sequence 276
Wide pairs 74
Wolf diagram 476

National Aeronautics and Space Administration
WASHINGTON, D. C. 20546

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

[illegible]
AIR FORCE WEAPONS LABORATORY (WLIL)
KIRTLAND AFB, NEW MEXICO 87117

ATT E. LOU BOSSARD, CHIEF, TECH. LIBRARY

NASA TT F-506
TT 68-50307

Printed in Jerusalem, Israel